The Evolution of Neurotactin and other CLAM Family Proteins

Honors Research Thesis

Presented in Partial Fulfillment of the Requirements for Graduation

"with Honors Research Distinction in Molecular Genetics" in the Undergraduate Colleges of

The Ohio State University

By

Suchita Sampath

The Ohio State University

April 2014

Project Advisor: Dr. Mark Seeger, Department of Molecular Genetics

INTRODUCTION:

Embryogenesis is a process consisting of cell division and cellular differentiation in a developing organism. It is during this process that drastic morphogenetic events occur which lead to a specific spatial distribution of cells. The morphogenetic events are regulated by a variety of cellular movements that range from changes in cell shape to segregation of cells based on the type, cell rearrangements as well as cell migration (1). Most of these movements are mediated by the cell's ability to selectively adhere to other cells. Specialized molecules, known collectively as adhesion molecules, have a complex, dynamic, spatio-temporal dependent expression pattern during the organism's development. Their complex expression pattern enables a cell to bind selectively to other cells (2). The function of adhesion molecules is also thought to go beyond simple homophilic or heterophilic cell to cell adhesion, in that the contact between two cells triggers intracellular signaling through the cytoplasmic domains of some cell adhesion molecules (1). Adhesion molecules are, therefore, crucial during development and are particularly so in the development of the nervous system, as they are instrumental to the mechanism controlling neurite growth, growth cone guidance, axon fasciculation and synaptic target recognition (3).

The cholinesterase-like adhesion molecules (CLAMS) protein family is a class of heterophilic cell adhesion proteins that are found in a number of metazoa and are involved in the development of multiple tissues (4). The members of this protein family share a number of different features besides the similarity in function. The majority of the proteins in the CLAMS family are transmembrane proteins with an extracellular cholinesterase-like domain that is conserved. However, all of the members of the family lack one or more of the residues necessary for the catalytic triad to function. It therefore appears that the enzymatic function of the

cholinesterase-like domain is not necessary but does play a role in the cell's ability for heterophilic adhesion (4,5).

There are several members of the CLAM family present in *D. melanogaster*, including Gliotactin (Gli), Neuroligin 1 (Nlg1), Neuroligin 2 (Nlg2), Glutactin (Glt), and Neurotactin (Nrt). All of these proteins encode the characteristic extracellular cholinesterase-like domain of the protein family. Nrt, however, is the only type II transmembrane protein meaning that the protein is oriented with the carboxyl portion of the protein on the extracellular face of the membrane (4). All of the other transmembrane CLAM proteins have the cholinesterase- like domain of the protein on the extracellular face (See Figure 1).

Figure 1. *D. melanogaster* CLAM Family Members

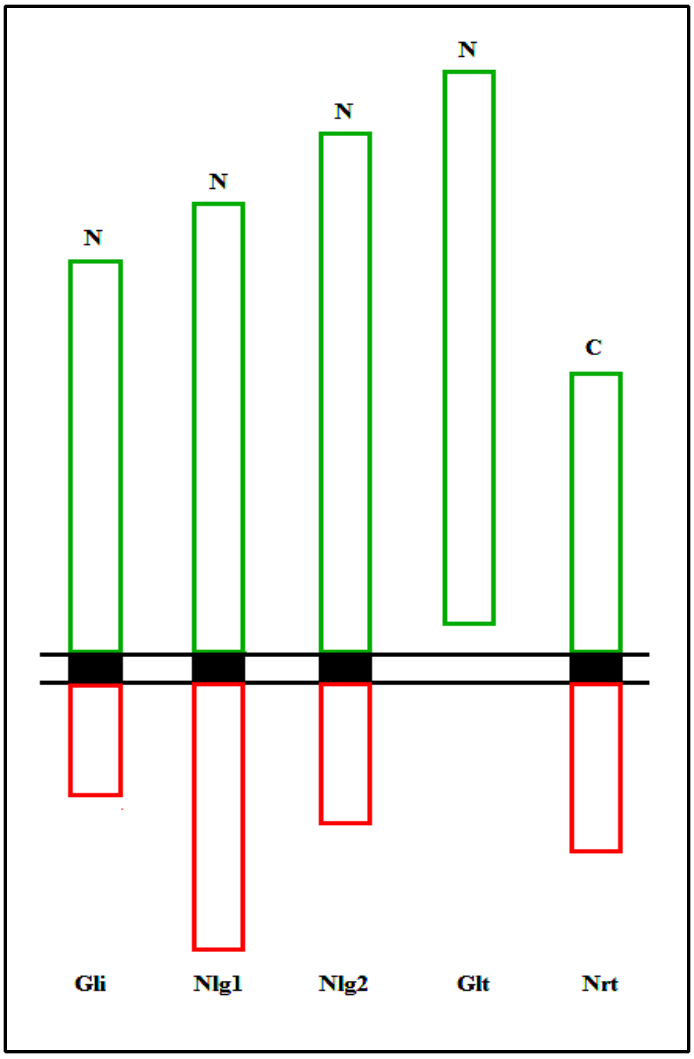


Figure 1. Diagram of the Cholinesterase-like Adhesion Molecules found in *D. melanogaster*. From Left to Right: Gliotactin (Gli), Neuroligin 1 (Nlg1), Neuroligin 2 (Nlg2), Glutactin (Glt), Neurotactin (Nrt). Gli, Nlg1, and Nlg2 are type I transmembrane proteins. Glt is a secreted protein. Nrt is the only type II transmembrane protein wherein the cholinesterase-like domain is flipped in comparison to the cholinesterase-like domain of the other CLAM family members. As a result, the C-terminus is facing the extracellular space in Nrt. In all of the other CLAM proteins the N-terminus is facing the extracellular space.

Gliotactin (Gli) is a single pass transmembrane protein and is expressed in a wide variety of invertebrate epithelial-like tissues (i.e. glia, epidermis, and the eye, wings and leg imaginal discs). Gli is localized to septate junctions where it, in turn, localizes other sepatate junction components such as Neurexin IV and Coracle. Gli, however, differs from these other septate junction proteins in that it is more specifically localized to the tricellular junction (a specific junction that meets at the convergence of septate junctions in three neighboring cells) (4). The septate junction, itself, forms a permeability barrier between cells and serves a similar function to tight junctions found in vertebrates. *Gliotactin* mutants result in defects in the permeability barriers which in turn lead to retarded maturation, a breakdown of the blood-nerve barrier and embryonic lethality (4). As is the case with CLAM family members, Gli serves as a heterophilic cell adhesion molecule but the corresponding ligand, however, remains unknown at this time (4). Gli is shown to be highly conserved in mosquito, honeybee and *C. elegans* as well as with the group of CLAM proteins known as Neuroligins (4).

In general, Neuroligins are type I transmembrane proteins which are present in diverse species (3 genes have been identified in mice, 5 genes in humans and homologues have been identified in *D. melanogaster* and *C. elegans*) (4). They were initially characterized as CNS specific binding partners for the beta-Neurexin family of transmembrane proteins but it was later noted that Neuroligins are expressed over a wide range of developmental periods as well as in a wide variety of tissues outside of the CNS. The latter observation, in turn, suggests that Neuroligins have non-Neurexin ligands as well (4). Rat and mouse Neuroligin 1 (Nlg1) seems to be expressed specifically in the CNS and are concentrated at synaptic junctions. It has been shown to induce presynaptic differentiation and play a role in the formation of inhibitory synapses (4). Rat Neuroligin 2 (Nlg2) is also expressed in the CNS but is localized primarily in

the inhibitory synapses *in vivo* and has, like Nlg1, been demonstrated to induce presynaptic differentiation. Much less is known about Nlg2 expression outside of the CNS but has been found to be expressed in a broader range of tissue types compared to the other Neuroligins (4).

Glutactin (Glt) is another CLAM family member that is highly expressed in the CNS but whose function is not completely known. Glt encodes an acidic sulfacted glycoprotein that, unlike other CLAM proteins, is secreted and localized to the basement membranes in *Drosophila* embryos (4).

Neurotactin (Nrt), which encodes a type II transmembrane protein, has been identified, along with Glt, as one of the more divergent members of the CLAM family. Nrt has a 500 amino acid extracellular domain which is the cholinesterase-like domain and a 324 amino acid cytoplasmic domain. The protein is involved with heterophilic cell adhesion in *D. melanogaster* and is required for proper axon guidance in *D. melanogaster* development. Nrt is mainly found in neural tissue as well as mesoderm and imaginal disc (6). It is also only detected during cell proliferation and differentiation and is not present in adult tissue suggesting that its function is vital for proper development (1). The loss-of-function *Nrt* mutant phenotype showed defects in axon guidance in embryonic and post-embryonic development (3). Double mutant combinations of *Nrt* with other genes encoding adhesion/signaling molecules lead to synergistic phenotypes suggesting that there is a functional cooperation between adhesion/signaling molecules controlled by Nrt and the other adhesion/signaling molecule encoding genes (3).

Nrt expressing Schneider 2 (S2) cells do not aggregate suggesting that Nrt does not, by itself, mediate cell adhesion in a homophilic manner. However, the same Nrt transfected S2 cells bind to embryonic primary culture cells indicating that there is a Nrt ligand involved in the cell adhesion. It is also important to note that S2 cells that do not express Nrt do not bind to primary

culture cells. Nrt is, therefore, binding in a heterophilic manner to a ligand that is only expressed in a subpopulation of embryonic cells (7). Amalgam (Ama), a member of the immunoglobulin superfamily, encodes a 333 amino acid long secreted protein and is co-localized with Nrt throughout neuronal development (6). Ama has been identified as the corresponding ligand to Nrt and is required for Nrt- expressing cells to aggregate to themselves as well as with primary embryonic culture cells (6). Through aggregation assays performed with a series of truncated versions of Nrt it was determined that only the amino acids 347-482, which constitutes a section of Nrt extracellular domain, are necessary for Ama binding and cell adhesion. The cytoplasmic domain for Nrt, however, is required for Nrt-mediated cell-adhesion but not for Ama binding (6). The Ama-Nrt interaction has been shown to be involved with axon-pathfinding and a loss-of-function *Ama* mutant displayed a similar phenotype to the loss-of-function *Nrt* mutant phenotype (6). Previous results have also suggested that the Ama-Nrt interaction is involved in a signaling network involving Abelson tyrosine kinase which affects axon pathfinding, particularly in the growth cone (8).

Previous studies have shown that Ama is a relatively recently evolved protein as it is not present outside of the order Diptera and is thought to have arisen early in dipteran evolution as a result of a duplication of the Lachesin gene (9). The phylogeny of Neurotactin and its function in insects, other than *D. melanogaster*, is not well characterized. The aim of this project is to further characterize the evolution of the CLAM family proteins members Gliotactin, Neuroligin 1, Neuroligin 2, Glutactin and Neurotactin. A greater focus will be placed on Neurotactin and its phylogeny in dipteran species as well as in other insect orders.

MATERIALS AND METHODS

Basic Local Alignment Search Tool (BLAST) search:

Basic Local Alignment Search Tool (BLAST) is a bioinformatics algorithm that allows for the comparison of primary sequence information such as the nucleotides of DNA sequences or the amino acid sequences of different proteins. The program searches for significant patterns in the sequences and is used in conjunction with other algorithms to determine approximate sequence correlation. A BLAST search compares the query sequence with a database of sequences and correlates sequences within that database to the query sequence, yet with a certain degree of sensitivity (i.e. the database sequence resemblance to the query sequence needs to be above a certain threshold in order for the sequence to be included in the BLAST search results).

There are many specific programs within the BLAST search family of programs. The primary program utilize in this project was the tblastn search program which compares a protein sequence to the six-frame translations of a nucleotide database. This program is particularly useful in identifying similar proteins in unannotated nucleotide sequences, which was primary aim of this project. There are several BLAST search hosts available and many were used in this project: FlyBase, VectorBase, NCBI, BeetleBase, and FleaBase. Each host carries a database of protein sequences and DNA sequences for a specific order of organism or a series of organisms.

In order to determine if homologous proteins of the *D. melanogaster* CLAM family of proteins existed in other dipteran and non-dipteran species, BLAST searches were utilized. The amino acid sequences for Nrt, Glt, Gli, Nlg1 and Nlg2 have already been determined in *D. melanogaster* and it was with these sequences that BLAST searches were conducted. The amino acid sequences for each of the CLAM proteins were entered into the tblastn algorithm and "blasted" against the nucleotide Genome Assembly for the different species of interest. It is

important to note that in some cases the searches were conducted with the use of predicted transcripts but were confirmed with different programs and BLAST searches utilizing the genomic contig or scaffold.

Searches for homologous Nrt proteins were also conducted using the cytoplasmic domain of the sequence alone, rather than the entire amino acid sequence. The reason for searching with a truncated version of the protein was that the extracellular domain of the protein is the cholinesterase-like domain. This domain would cause other CLAM proteins to be included in the BLAST search results as the cholinesterase-like domain is conserved. The results of the BLAST searches done with just the full length Nrt protein sequence might contain sequences which are, in fact, more homologues with other CLAM members like Gli, Nlg1, Nlg2 and Glt for example. The cytoplasmic domain of Nrt is, therefore, the best marker to identify homologous proteins present in other species.

ClustalW:

To further confirm that the BLAST results were accurate, the full amino acid sequence of the potential homologous protein was entered into a ClustalW program provided by Biology Workbench which is a part of the San Diego Supercomputer Center. This program is a multiple sequence alignment program for DNA or proteins. For the purposes of this project the protein alignment was used. ClustalW not only provides an alignment of multiple sequences but also illustrates whether sequences are divergent from one another. The best matches for portions of the sequences are lined up so that similarities and differences can easily be visualized. The ClustalW program goes further to differentiate the degree of amino acid consensus in the alignments with symbols. Those amino acids which are fully conserved are marked with an asterisk below the alignment, conservation of strong amino acid groups are marked with a

semicolon below the alignment and lastly conservation of weak amino acid groups are marked with a single period below the alignment. Those amino acids that share no consensus across species have no demarcation.

The Biology Workbench ClustalW program also provides a cladogram of the sequences in question. This creates a visual representation of how closely related, or how divergent the two amino acid sequences are from one another.

Particularly close attention was paid to the cytoplasmic domain of the Nrt homologues as that was one feature used to determine whether the protein sequence, determined by the BLAST search, was indeed a homologue of Nrt. If significant correlation is present not only in the extracellular domain (the cholinesterase-like domain) but also in the cytoplasmic domain, then the certainty that the protein is a homologue of *D. melanogaster* Nrt increases.

Hydropathy Plot:

The ExPASY ProtScale program generates hydropathy plots of amino acid sequences. The hydropathy plot itself allows one to graphically represent the relative hydrophobic and hydrophilic nature of the amino acids that make of the protein sequence. Each amino acid is defined by a numeral value.

The amino acid scale is determined based on experimental studies wherein the peptides are determined to be hydrophilic or hydrophobic. For the purposes of this project the Kyte and Doolittle amino acid scale was used in the generation of the hydropathy plots. This scale will result in a graph where the peaks in the graphs (positive values) represent the extremely hydrophobic regions of the amino acid sequence.

The hydropathy plots prove to be particularly helpful in confirming whether the cholinesterase-like protein sequence discovered is actually Nrt. This is because Nrt is the only

type II transmembrane protein in the CLAM family of proteins. Unlike a type I transmembrane protein, a type II transmembrane protein does not have a signal sequence at the amino-terminus and the orientation will be reversed, such that the C-terminus will oriented extracelullarly and the N-terminus will be in the cytoplasm. Therefore, on a hydropathy plot there will only be one hydrophobic peak which is for the transmembrane domain. In a type I transmembrane protein (i.e. Gli, Nlg1 and Nlg2), there will be two hydrophobic peaks on a hydropathy plot. The two peaks correspond to the signal sequence and the transmembrane domain. Glt, however, is a secreted protein and will, therefore, will only have one hydrophobic peak which represents the signal sequence (See Figure 2).

Figure 2. *D. melanogaster* CLAM Family Member Hydropathy Plots

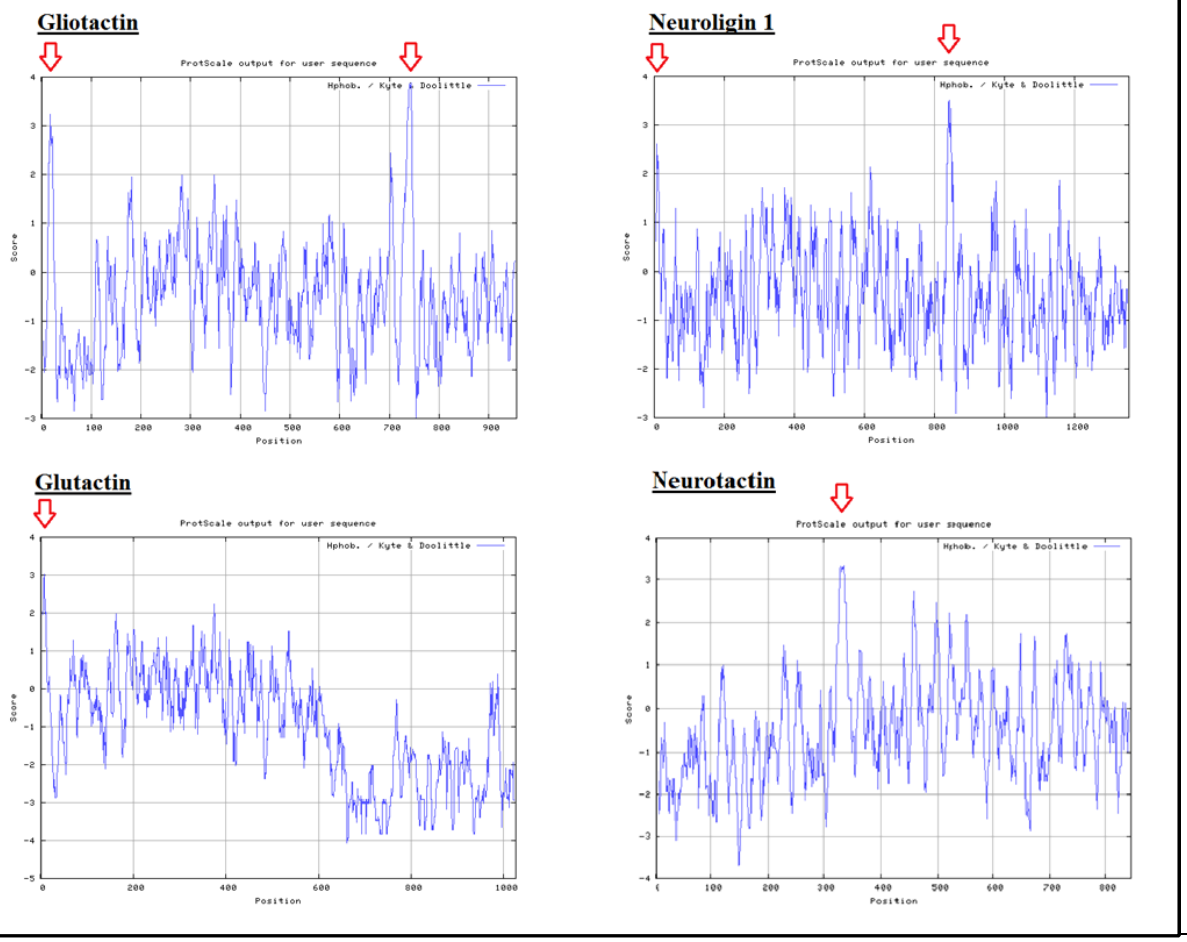


Figure 2. Hydropathy Plots of the *D. melanogaster* CLAM proteins: Gliotactin (Gli), Neuroligin 1 (Nlg1), Glutactin (Glt) and Neurotactin (Nrt). Gli and Nlg1 show attributes of a type I transmembrane protein with a two hydrophobic peak, one from the signal sequence and the second from the transmembrane domain. Glt only have one hydrophobic peak corresponding to its signal sequence. There is not a second hydrophobic peak related to a transmembrane domain as Glt is a secreted protein. Nrt also has one hydrophobic peak but the peak corresponds to its transmembrane domain. There is not a hydrophobic for the signal sequence because Nrt is a type II transmembrane protein.

Transmembrane Prediction Plot:

The TMpred Server hosted my ExPASy makes predictions of membrane-spanning amino acid sequences and their orientation. The program's algorithm is based on a database of naturally occurring transmembrane proteins and the prediction is made by using a scoring system that was established using the database of characterized transmembrane proteins. The results from this program should mimic those found in the hydropathy plot as the transmembrane domain should be the most hydrophobic region of the protein sequence.

The transmembrane prediction program also provides information on the possible orientation of the amino acid sequence. The prediction is determined by judging the varying charges of the amino acids. The charges and grouping of charged amino acids impact whether the N-terminus or the C-terminus is predicted to reside in the cytoplasm or in the extracellular space. The prediction of orientation can help provide further support that the homologue determined (particularly in the case of Nrt) is indeed a homologue of the CLAM protein in question. This proves to be useful in characterizing the homologues of Nrt because Nrt is the only CLAM family protein which has the C-terminus in the extracellular space (a feature of being a type II transmembrane protein). All other CLAM proteins are oriented with the N-terminus in the extracellular domain.

RESULTS AND DISCUSSION

The evolutionary characterization of the CLAM family members: Gliotactin, Neuroligin 1, Neuroligin 2, Glutactin and Neurotactin, was conducted through the utilization of BLAST searches in conjunction with ClustalW alignments. Through these programs several homologues for each CLAM protein were determined. For Gli, homologues were identified in *D. melanogaster, D. ananassae, D. grimshawi, D. pseudoobscura, D. virilis, D. willistoni, C. capitata, L. longipalpis, P. papatasi, A. aegypti, A. gambiae, C. quinquefasciastus, T. castaneum* and *P. humanus* (See Table 1, Figure 3 and Figure 4). Nlg1 was highly conserved in *D. melanogaster, D. ananassae, D. grimshawi, D. pseudoobscura, D. virilis, D. willistoni, L. longipalpis, P. papatasi, A. aegypti, A. gambiae, C. quinquefasciastus* and *D. pulex* (See Table 2, Figure 5 and Figure 6). In Nlg2, homologues were found in *D. melanogaster, D. ananassae, D. grimshawi, D. pseudoobscura, D. virilis, D. willistoni, C. capitata, A. aegypti, A. gambiae,* and *C. quinquefasciastus* (See Table 3, Figure 7 and Figure 8). Homologues for Glt were identified in *D. melanogaster, D. ananassae, D. grimshawi, D. pseudoobscura, D. virilis, D. willistoni,* and *R. prolixus* (See Table 4, Figure 9 and Figure 10). Lastly, Nrt was shown to be highly conserved in *D. melanogaster, D. ananassae, D. grimshawi, D. pseudoobscura, D. virilis, D. willistoni, C. capitata, M. domestica, G. morsitans, A. aegypti, A. gambiae, C. quinquefasciastus, T. castaneum, N. vitripennis, P. humanus,* and *R. prolixus* (See Table 5, Figure 11 and Figure 12).

The process to determine and confirm the homologues in the CLAM proteins was consistent across. Neurotactin was the focus of this project and 15 homologues, outside of the *D. melanogaster* Nrt, were identified. Further confirmation of the identity of the homologues was conducted through the use of the Transmembrane Prediction plots and the Hydrophobicity plots (See Figures 1a.- 1p.). The latter, in particular, shows a peak of hydrophobicity towards the

beginning which is the characteristic structure of a type II transmembrane protein. As Nrt is the only type II transmembrane protein in the CLAM family of proteins, one can say with a decent amount of certainty that the protein sequences found are indeed the homologues to Nrt.

One observation in the Transmembrane Prediction Plots to address is that the Prediction Plot suggests possible orientations for the protein (i.e. whether the N-terminus is oriented towards the cytoplasm or towards the extracellular space). As is evident there is no real consistency between orientation prediction, the score of the orientation prediction and how closely related the organisms are to *D. melanogaster*. For example only three species besides *D. melanogaster* are predicted with N-terminus on the inside (*D. grimshawi, M. domestica,* and *P. humanus*) whereas all of the other homologues are predicted to have the N-terminus on the outside. There is no consistency between how related the species are to one another, furthermore, the scores that accompany the prediction are similar across the board (See Figure 3). A lot of weight, as a result, cannot be placed upon the orientation predictions suggested by the Transmembrane Prediction Plot.

A similar procedure was conducted with the other CLAM family proteins: Gli, Nlg1, Nlg2, and Glu to determine homologues in species outside of *D. melanogaster*.

Based on the alignments provided by the ClustalW program we were able to determine that there are certain sequences that are conserved across all species. It is expected that many of these conserved sequences will occur in the cholinesterase-like domain as the presence of this domain is linked to the protein's function in adhesion but many conserved sequences occurred in the cytoplasmic domain. For the cytoplasmic domain of NRT there was multiple instances of conserved sequences, even in species whose divergence from one another occurred millions of years ago (See Homology list for highlighted conserved sequences).

With the CLAM members Gli, Nlg1, Nlg2 and Glt, homologues were difficult to identify as many of the suggested homologues included the conserved cholinesterase-like domain but outside of that domain there was very little conservation. Also the predicted homologues through the BLAST resulted in proteins that were significantly shorter in amino acid length than the *D. melanogaster* counterpart. There are multiple reasons as to why the predicted protein sequences were significantly shorter than the *D. melanogaster* counterpart. 1. The predicted protein had a cholinesterase-like domain which correlated to the cholinesterase-like domain in the CLAM protein being searched for. It was therefore suggested as a possible homologue to the CLAM being searched for but outside of that domain, there was no conservation of sequence. 2. There is a possibility of mutation occurring and certain sections of the protein sequence being excised through evolution. This option is the less parsimonious option. The probability that no other portion of the CLAM protein sequence, outside of the cholinesterase-like domain, is conserved is unlikely to be a true homologue to the CLAM protein in question. This is further supported by the fact that homologues were discovered in species that were distantly related to one another and had stretches of sequence conservation outside of the cholinesterase-like domain (See respective Homologue List).

An example of this is Glutactin, where many homologues were not determined. The BLAST searches suggested many homologue possibilities for Glutactin but when they were entered into the ClustalW program, very few showed homology. There were possible homologues in species more closely related to *D. melanogaster* but when the homologue for Glt was characterized in *R. prolixus* (the most distantly related insect species searched for) it was clear from the alignment that there was significant homology present in the *R. prolixus* Glt

homologue. It can therefore be inferred that the homologues that are more closely related to *D. melanogaster* should show a greater amount of conservation, which was not the case.

The most thorough characterization was conducted with Neurotactin. The most distantly related species in which a homologue was characterized was *R. prolixus* (the kissing bug). Due to how distantly related *R. prolixus* is from *D. melanogaster* attempts were made to find a Neurotactin homologue in non-insect species: *Daphnia pulex* (the water flea) and *Ixodes scapularis* (the deer tick). Unfortunately no clear homologue to Neurotactin was determined in either non-insect species. It was evident through the search, however, that there are many cholinesterase-like domain proteins present in non-insect species and one protein in *D. pulex* proved to be a homologous to *D. melanogaster* Nlg1.

Attempts to find a homologue of Nrt outside of the class *Insecta* were conducted to try and shed further light onto the emergence of Nrt. As this was unsuccessful, the emergence of Nrt can only be traced back to divergence of insects. At some point in time, a CLAM protein, similar to type I transmembrane proteins seen in Figure 1 underwent a mutational event such that it lost its signal sequence and was reorientated such that the C-terminus was in the extracellular space. A transmembrane domain and a cytoplasmic domain were also fused upstream of the cholinesterase-like domain. It is still unclear whether the mutation occurred before the divergence of insects, since a Nrt homologue has not yet been identified in a non-insect species. Based on the homologues identified in this project, Nrt can only be traced as far back as the divergence of holometabolous and hemimetabolous.

Table 1. Gliotactin Homologues

D. melanogaster	MMHKLKYRDKLKWLLALLVLIGTCFIQTRGQTRDPRFYSRPGVDYHWPNPGDPD YRTYTFNDRRYGHYQPNGYGANYPGRNPPGQYPQGMPNEDRFRFDPNDPNARTQ FPGVLAGWREDLQGKQRRDSLTLERDVFVTTNYGQVQGFKVYMYDNPDPKSFYR PYHSTVDRVMGECSVFLGIPYALPPTFEGRFKPPRVHRGWQLLQAVDFGPACPQPV RYTGATKGIMDMDEDCLYLNVYSPKTGAGVAQKYPVMVYIHGGEFIRGASNLFQ GHILASFYDVVVVTLNYRLGALGFLSTGDENSPGNYGILDQAMALRWVYDNIEFF NGDRNSITLFGPGAGGASAGLLMVAPQTRNIVRRVIAQSGSALADWALIQDKYRA QNTSRVLGQLLGCSIESSWKLVNCLRTGRSFYELGNAEFSPQVGSFPWGPVLDHNF TLPGDDWYEGWREKDWRFLTQTPETLIRAGKFNRNIQYMTGVTTQEAAFFVAQN ESLSPYYELDGRFFDQKIREHVFRYNYTLNPNGVYEAIKYIYTFWPDPNNNTIIRDQ YINMLSDLYYRAPVDQMVKLMLEQKVPVYMYVLNTTVEALNLPQWRKYPHDIER YFLTGAPFMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYTNFARYGNPTPQQV LGMHFQRAYQGEIRYLNINTTYNSSILLNYRQTECAFWTQYLPTVIGVLVPTYPPTT EYWWEPKEPLQIAFWSMSVACFFLIVLVVICCIMWRNAKRQSDRFYDEDVFINGE GLEPEQDTRGVDNAHMVTNHHALRSRDNIYEYRDSPSTKTLASKAHTDTTSLRSPS SLAMTQKSSSQASLKSGISLKETNGHLVKQSERAATPRSQQNGSTAKVASPPVEEK RLLQPLSSTPVTQLQAEPAKRVPTAASVSGSSRSTTPVPSARSTTTHTTTATLSSQPA AQPRRTHLVEGVPQTSV
D. ananassae	MMHKLKYRDKLKWLQALLICVCLSHFAEAQTRDPRFYSRPGVDWPNPGDPDYRT YTFNDRRYGHYQPNGFGANYPGRNSPGQYPQGMPNEDRFRFDPNDPNARTQFPG VLAGWREDLQGKQRRDSLTLERDVFVTTNYGQVQGFKVYMYDNPDPKSFYRPYH STVDRVMGECSAFLGIPYALPPTFEGRFKPPRIHRGWQLLQAVDFGPACPQPVRYT GATKGVMDMDEDCLYLNVYSPKTGAGVAQKYPVMVYIHGGEFIRGASNLFQGHI LASFYDVVVVTLNYRLGALGFLSTGDENSPGNYGILDQAMALRWIHDNIEFFNGD RDSITLFGPGAGGASAGLLMVAPQTRNIVKRVIAQSGSALADWALIQDKYRAQNT SRVLGQLLGCSIESSWKLVNCLRTGRSFYELGNAEFAPQVGSFPWGPVLDHNFTVP GDDWYEGWREKDWRFLTQTPETLIRAGRFNRNIQYMTGVTTQEAAFLVAQNESL SPYYELDGRFFEQKIREHVFRYNYTLNPNGVYEAIKYMYTYWPDPNNNTIIRDQYI NLLSDLYYRAPVDQMVKLLLEQKVPVYMYVLNTTVEALNLPQWRKYPHDIERYF LTGAPFMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYSNFARYGNPTPQQVLG MHFQRAYQGEIRYLNVNTTFNSSILLNYRQTECAFWTQYLPTVIGVLVPTYPPTTE YWWEPKEPLQIAFWSMSVTCFFLIVLVVICCIMWRNAKRQSDRFYDEDVFINGEG MEPEPDARGVDNAHMVTNHHALRSRDNIYEYRDSPSSKTLASKAQTDTTSLRSPSS LAMTQKSGSQASLKSGISLKETNGSLTKSVSARSQANGNGTTAKSAAAAPVEEKR LLQPLSSTPVPQLQAEPAKRVPTAASVSGSSRSTTPVPSARTTTTTTTATLSSQPAPQ PRRTQLVEGVPQTSV
D. grimshawi	MSEQQLQLATMSQKFKYRDKFECLSGILMVLLVLQPLLPHIDAQTRDPRFYSRVGV DYQWPNPGDPEYRTYTFNDRRYGHYQPNGYGANYPGRNPPGQYPQGMPNEDRFR FDPNDPNARTQFPGVLAGWREDLQGKQRRDSLTLERDVFVTTNYGQVQGFKVYM YDNPDPKSFYRPYHSTVDRVMGECSTFLGIPYALPPTFEGRFKPPRLHRGWQLLQA VDFGPACPQPVRYTGATKGVMDMDEDCLYLNVYSPKTGAGVAQKYPVMVYLHG GEFVRGASNLFQGHILASFYDVVVVTLNYRLGALGFLSTADENSPGNYGILDQAM ALKWVYDNIEFFNGDRESITLFGPGAGAASAGLLMVAPQTRNIVKRVIAQSGSALA DWAFIQDKYRAQNTSRVLGQVLGCSMDSSWKLVNCLRTGRSFYELGNAEFSPQV GSFPWGPVLDHNFTLPGDDWYEGWREKDWRFLTQTPETLIRAGRFNRNIQYMTG VTTQEAAFFVAQNESLAPYYELDGRFFEQKVREHVFRYNYTLNPNGVYEAIKYMY TFWPDPNNNTIIRDQYINMLSDLYYRAPVDQMVKLLLEQKVPVFMYVLNTTVEAL NLPQWRKYPHDTEHYFLTGAPFMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQT YSNFARYGNPTPQQVLGMHFQRAYQGEIRYLNINTTYNSSILLNYRQTESAFWTQY LPTVIGVLVPTYPPTTEYWWEPKEPLQIAFWSMSVACFFLIVLVVICCIMWRNAKR QSDRFYDEDVFINGEGMEPEPDPRGVDNAHMVTNHHAMRSRDNIYEYRDSPSTKT

	LASKVHTDTTSLRSPSSLAMTQKSSSQASIKSGISLKETNGSLVKSERAATPRSQVN GTVSSKAPPVEEKRLLQPISSTPVAQLQSEPTKRTPTASSVASSSRSTTPVPSARTTTT TATLTTQPAAQPRRTQLVEGVPQTSV
D. pseudoobscura	MMHKLKYRDKSKWLTAILFLVVASLGPLVGGQTRDPRFYSRPGVDYHWPNPGDP DYRTYTFNDRRYGHYQPNGFGANYPGRNPPGQYPQGMPNEDRFRFDPNDPNART QFPGVLAGWREDLQGKQRRDSLTLDRDVFVTTNYGQVQGFKVYMYDNPDPKSFY RPYHSTVDRVMGECSAFLGIPYALPPTFEGRFKPPRAHRGWQLLQAVDFGPACPQP VRYTGATKGVMDMDEDCLYLNVYSPKTGAGVAQKYPVMVYIHGGEFIRGASNLF QGHILASFYDVVVVTLNYRLGALGFLSTGDENSPGNYGILDQAMALRWIHDNIEFF NGDRDSITLFGPGAGGASAGLLMVAPQTRHMVKRVIAQSGSALADWALIQDKYR AQNTSRVLGQLLGCSIESSWKLVNCLRTGRSFYELGNAEFSPQVGSFPWGPVLDHN FTLPGDDWYEGWREKDWRFLTQTPETLIRAGKFNRNIQYMTGVTTQEAAFLVAQ NESLGPYYELDGRFFEQKIREHVFRYNYTLNPNGVYEAIKYMYTYWPDPNNNTIIR DQYINMLSDLYYRAPVDQMVKLLLEQKVPVFMYVLNTTVEALNLPQWRKYPHDT ERYFLTGAPFMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYSNFARYGNPTPQ QVLGMHFQRAYQGEIRYLNVNTTFNSSILLNYRQTECAFWTQYLPTVIGILVPTYPP TTEYWWEPKEPLQIAFWSMSVTCFFLIVLVVICCIMWRNAKRQSDRFYDEDVFING EGLEPEPDPRGVDNAHMVTNHHAMRSRDNIYEYRDSPSSKTLASKAHTDTTSLRSP SSLATTQKSGSQASLKSGISLKETNGSLVKSERAATPRSQPNGTTAKAGPVEEKRLL HPLSSTPVPQLQSEPTKRAPASVSGSSRSTTPVPSARTTTTTATLSSQPAAQPRRTQL VEGVPQTSV
D. virilis	MAQKFKYRDKLKCLLAMLMALMLPQSIDAQTRDPRFYSRVGVDYQWPNPGDPD YRTYTFNDRRYGHYQPNGYGANYPGRNPPGQYPQGMPNEDRFRFDPNDPNARTQ FPGVLAGWREDLQGKQRRDSLTLERDVFVTTNYGQVQGFKVYMYDNPDPKSFYR PYHSTVDRVMGECSVFLGIPYALPPTFEGRFKPPRLHRGWQLLQAVDFGPACPQPV RYTGATKGVMDMDEDCLYLNVYSPKTGAGVAQKYPVMVYIHGGEFVHGASNLF PGHVLASFYDVVVVTLNYRLGALGFLSTGDENSPGNYGILDQSMALKWVHDNIEF FNGDRDSITLFGPGAGGASAGLLMVAPQTRNIVKRVIAQSGSALADWALIQDKYR AQNTSRVLGQVLGCSIDSSWKLVNCLRTGRSFYELGNAEFPPQVGSFPWGPVLDH NFTLPGDDWYEGWREKDWRFLTQTPETLIRSGRFNRNIQYMTGVTTQEAAFFVAQ NESLAPYYELDNRFFEQKVREHVFRYNYTLNPNGVYEAIKYMYTYWPDPNNNTII RDQYINMLSDLYYRAPVDQMVKLLLEQKIPVYMYVLNTTVEALNLPQWRKYPHD TEHYFLTGAPFMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYSNFARYGNPTP QQVLGMHFQRAYQGEIRYLNINTTYNSSILLNYRQTECAFWTQYLPTVIGVLVPTY PPTTEYWWEPKEPLQIAFWSMSVACFFLIVLVVICCIMWRNAKRQSDRFYDEDVFI NGEGMEPEPDARGVDNAHMVTNHHALRSRDNIYEYRDSPSTKTLASKAQTDTTSL RSPSSLAMTQKSGSQASLKSGISLKETNGSLVKSERAATPRSQTNGTSAKAAPIEEK RLLQPLSSTPVAQLQAEPTKRMAAAPSVASSSRSTTPVPSARTTTTTATLTAQPAVQ PRRTQLVEGVPQTSV
D. willistoni	MLHKLKYRDKLKCSYAICLILILSFLATLTDGQTRDPRFYSRVGVNDYHWPNPGDP DYRTYTFNDRRYGHYQPNGFGANFPGRNPPGQYPQGMPNEDRFRFDPNDPNART QFPGVLAGWREDLQGKQRRDSLTLERDVFVTTNYGQVQGFKVYMYDNPDPKSFY RPYHSTVDRVMGECSVFLGIPYALPPTFEGRFKPPRPHRGWQLLQAVDFGPACPQP VRYTGATKGVMDMDEDCLYLNVYSPKTGAGVAQKYPVMVYIHGGEFIRGASNLF QGHMLASFYDVVVVTLNYRLGALGFLSTGDENSPGNYGILDQALALKWVYDNIEF FNGDRESITLFGPGAGGASAGLLMVAPQTRNIVKRVIAQSGSALADWALIQDKYR AQNTSRVLGQLLGCSIESSWKLVNCLRTGRSFYELGNAEFSPQVGSFPWAPVLDHN FTLPGDDWYEGWRQKDWRFLTQTPETLIRAGRFNRNIQYMTGVTTQEAAFFIAQN ESLAPYYEIDGRFFEQKVREHVFRYNYTLNPNGVYEAIKYMYTFWPDPNNNTIMR DEYINMLSDLYYRAPVDQMVKLMLEQKVPVYMYVLNTTVEALNLPQWRKYPHD TEHYFLTGAPFMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYSNFARYGNPTP QQVLGMHFQRAYQGEIRYLNINTTYNSSILLNYRQTECAFWTQYLPTVIGVLVPTY

	PPTTEYWWEPKEPLQIAFWSMSVTCFFLIVLVVICCIMWRNAKRQSDRFYDEDVFI NGDGIEPETDPRGVDNAHMVTNHHALRSRDNIYEYRDSPSSKTLASKAHTDTTSLR SPSSLAMTQKSGSQASLKSGISLKETNGSLVKSERAATPRSQTNGKSPNPGPIEEKR LLQPISSTPVPQLQAEPVKRTTPGPAASISGSSRSNTPVPMARTTTTTTATLTTQPPA AGRSVDQPRRTELVEGIPQTSV
Ceratitis capitata	MVHRKVVQLINTRPLRMNTFQYHDKVKSAVMLVIASYLVLSNLICVKAQTRDPRF FSRPGVNDYNWPNPGDPDYRTYIFNDRRYGHYLPNGYGTNYPGRNSPGQYPQGM PNEDRFRFDPNNPNAAHTPFPGILAGWREDLQGKQRRDSLTLDRDVFVTTNYGQV QGFKVYMYDNPDPKSFYRPYHSTVDRVMGECSVFLGIPYALPPTFEGRFKPPRLHR GWQLLQAVDYGPACPQPVRYTGATKGIMDMDEDCLYLNIFSPKTGAGVAQKYPV MVYIHGGEFIRGASNLFQGHILASFYDVVVVTINYRLGALGFLSTGDDNSPGNYGIL DQVMALKWIYDNVEFFNGDRESITLFGPGAGAASAGLLMVAPQTKNIVKRVIAQS GAAMADWALIQDKYRAQNTSRVLGQLLGCSIESSWKLVNCLRTGRSFYELGNAEF PPQVGTFPWGPVLDHNFTVPGDDWYEGWREKDWRFLTQTPEYLIRQGHFNRGLQ YMTGVTTQEASFFVAQNESLAPYYEIDHTFFDAKIREHVFRYNYTLNPNGVYEAIK YMYTYWPDPNNSSVIRDQYINLLSDLYYRAPVDKMVKLLLEQRIPVYMYVLNTTV EALNYPQWRKYPHNIEHYFLTGAPFMDIEFFPKKEHLQRNMWTDNDRNMSHFFM QTYSNFARYGNPTPQQVLGLHFQRAYQGELRYLNINTTFNSSILLNYRQTECAFWT QYLPTVIGVLVPTYPPTTEFWWEPKEPLQIAFWTMSVACFFLIVLVVICCIMWRNA KRRSDFYDDDVFINPDGAELPDEVHTGVDNTHMVANPHAMRSRDNIYEYRDSPSS KTLASKIHTDTTSIRSPSSLAMTQKSGSQSSLKSTMSLKETTGGTLTSCSSSARPSRT NGLHNGSAATIQTPKVFEEKRLITRELTKPEPISSTPLVTTSTAKISITSEPISLKPTSNI IATSTSPNLTGTVSSRSTTPVPSARTHTHTTTATVTSAPQVQRPVLPTSSAQAQSRA QARTQVIEGVPQTSV
Lutzomyia longipalpis	MAQDLRQSAQNSKTKPQKRISVRWRGLAVIVGVMCVLPVGIRAQWDQYTTRDPR FYSRDGVNNYNPPNPGDPEYRTYTYNDRRYGYYQPAGYGHLYPGQKLPGQYPNP HDTPLGDDRFKFDPSDPNTFQTPFPGVLGGWREDLQGKLRRDSLTLDRDVFVTTN YGQVQGFKVYLYDNPDPLSFYRPWHSTVDRIMGQCSVFLGIPYALPPTHEGRFKPP RPHRGWQLMQAVDFGPACPQPVRYTGATKGIRDMDEDCLYLNVFSPKTGSGVPQ KYPVMVYIHGGEFTRGASNLFPAHVLASFYEVVVVSINYRLATAEWALTVDKYRA QNTSIVYGKLLGCYIESSYRLVECLRLGRSFYEIANAEFAPDVGFFPWGPVLDMNL TYPGDNWYEGWREKDWQFLKETPEQMIKKGLFNKGLHYMTGVTTQEAAYVLYQ NESLAPYFDIDEDFVDQKIRELVLRYNYTLNLNGTYEAIKYMYTHWPDPSNKTYIR EEYINLLSDFIYRAPSDQMTKLLVEQNVPVYMYVLNTTVEGMRLQNWRAVPHNIE HYFLTGAPFMDVEFFPKRDRLERNMWTDNDRNMSHFFMKTFSDFARFGNPTPQQ VLGLHFDKAINGELRYLNLNTTFNSTIRMNYRQKQCAFWTQYLPTVVGVLVPTYP PSTEYWWEPQQPLQIAFWSISAVCLFLVILLVVCCILWRNAKRQSDRFYDGDVFMV DNTGADGDAGIENISQTHPLRSRENIYEFRDTPKKSGPPTDISSVRSPSSLATNPTAK SGSQSSLKSAISLKEGAQNGRDGNYDRNLNNKGKGSDGKNLLADFEKSREAMTSS PVDVKTSSHGIPIPASRKTVKTSKTQLVEGVPQTEV
Phlebotomus papatasi	MAIVRDLRQSVQNSLPKPQVLKSERWRGLLVLVGVICVLPMGIRAQWDQYTTRDP RFYSRDGVYNYNPPNPGDPEYRTYTYNDRRYGYYQPAGYGHLYPGQKLPGQYPN PHDTPLGDDRFKFDPSDPNTFQTPFPGVLGGWREDLQGKLRRDSLQLDRDVFVTT NYGQVQGFKVYLYDNPDPLSFYRPWHSTVDRIMGQCSVFLGIPYALPPTLEGRFKP PRPHRGWQLMQAVDYGPACPQPVRYTGATKGIRDMDEDCLYLNVFSPKTGSGVP QKYPVMVYIHGGEFMKGASNLFPAHVLASFYDVVVVSINYRLGALGFLSTGDENS PGNYGILDQAMAIRWVYDNIEAFNGDRHSITLFGPGAGAASAGLLMVAPQTRNIV TRVIAQSGSATADWALIVDKYRAQNTSIVYGKLLGCYIESSYRLVECLRLGRSFYEI ANAEFAPDVGFFPWGPVLDMNFTYPGDNWYEGWREKDWQFLKETPEQLIKKGM FNKGLHYMTGVTTQEAAYAIYQNESLAPYFEIDEEFFDQKVRELTLRYNYTLNLN GTYEAIRYMYTHWPDPKNKTFIREEYINMLSDFIYRAPSDQMTKLLVEQNVPVYM YVLNTTVEGLRLQDWRAVPHNAEHYFLTGAPFMDVEFFPKRDRLDRNMWTDND

	RNMSHFFMKTFSDFARFGNPTPQQVLGLHFDKAINGELRYLNLNTTFNSTIRMNYR QKQCAFWTQYIPTVVGVLLPTYPPSTEQSDRFYDGDVFMVDNTGADGDGGGIENI SQTHPLRSRENIYEYRDSPTKKSGPPTDISSIRSPSSLATTQISAKTGSQSSLKSAMSL KEGAQNGKDSTYDRNLNNKGKGSDSRNLLADFEKSREAMTSSPVDGVKTAGTPIP ATRKVVKTSKTQLVEGVPQTEV
Aedes aegypti	MALNTNCGAAKRYQQAKHSAKTTKSSDGGGGCSVRKKCVHLSSSTTGSEWIHAK INVLFLCFAVFISLVDGQYHTTRDPRWYSREGDFSYHLPNPGDPDYRTYTYNNRRY GYYQPNGYGQQYPGQNLPGQYPQNTGLGEDRFKYDPTNPNNVQTPFPGVLGGWR EDLQGKLRRDSIQLDRDVFVTTNYGQVQGFKVHLYDNPDPKSFFRPWHSPVDRV MGTCSTFLGIPYALPPTFEGRFKPPRPHRGWQLLQAVDYGPACPQPVQYTGATKGI RDMDEDCLYLNVFSPNTQSGVAQRFPVMIYIHGGEFVRGASNVFPGHMLAAFYEV VVVTFNYRLGALGFLSTGDENSPGNYGILDQIMAVRWVYDNIEAFNGDRNSITLFG PGAGAASAGLLMVAPQTKDIVTRVIAQSGAAVADWALIVDKYRAQNTSRVFGQM VGCSIETSWKLVNCMRQGRSFFELGNAEFAPHVGLFPWGPVLDINFTFPGDEWYE GWRERDWHFLAETPEELIRKGHYNRGLHYMSGVTLQEAAFVISQNESLAPYFEVD QKFFEQKVWELVYRYNYTLNLNGTYEAIKYMYTHWPDPRNTTYIREQYINLLSDF LYRAPADKMTKLLVERRVPVYSYVMNTTIEGLKLPEWRKVPHDIEHYLLTGAPFM DVEFFPRKARLDRLMWTDNDRNMSHFFMKTFSDFARYGNPTPQQVLGLFFEPARN GELKYLNLNTTYNSSIRMNYRQTESAFWTQYIPTVVGVMVPTYPPTTEFWWEPKE PLQIAFWSMSAACLLLIVMVVICCILWRNAKRQSDRFYDDGVIVVDNPESERDVGI DNNVLQAQPLRTRDNIYEYRDSPTKYKSPPTDATSIRAPSSLAMTQTTAKSISGSQS SLRSAISMKESTNISPIPRSESNFDRNIKYAERGVDTEINKNGKYPRTDSRGALSDTK SREFSASTPIEPAPRSQQPSSQPASRGASRAGYDRSNNASRTNLIEGIPQTEV
Anopheles gambiae	MTLSREHHHHASLLLLVVIVGVASLVGVCDGQYYTTRDPRWYSREGDFNYKIPNP GDPEYRTYTFNNRRYGYYQPNGYGQQYPGQSLPGQYPSNTGLGDDRFKYDPTNPT LAQTPFPGVLGGWREDLQGKLRRDSMQLERDVFVTTNYGQVQGFMVYLYDNPDP KNLYRPWHSNVDRIMGKCSTFLGIPYALPPTHEGRFKPPRQHRGWQLVQAVDYGP ACPQPVQYTGATKGIRDMDEDCLYLNIFSPNTQAGVAQRYPVMIYIHGGEFVRGA SNTFPGHMLAAFYDVVVVTFNYRLGALGFLSTGDENSPGNYGILDQIMAVRWVYE NIESFNGDRNSITLFGPGAGGASAGLLMVAPQTKDIVTRVIGQSGSALADWALIVD KWRAQNTSRVFAQNVGCSIETSWKMVNCMRNGRSSYELGNAEFPPHVGLFPWGP VLENNFTFPGDSWYEGWSERDWHFLAETPESLIRRRHFNRGLHYMSGVTLQEAAF VITGNKSLAPYFEVDERFFDQKVRELVLRYNYTLNLNGTYEAIKYMYTYWPDPKN TTFIREQYINLLSDFLYRAPTDKLIKLLVEQNVPVYSYVMNTTIEGLKLAEWCKVPH DIEHYLLAGAPFMDVEFFPRRERFDRLMWTDNDRNMSHFFMKTYSDFARHGNPTP TQVLGLYFEKARNGELKYLNLNTTYNSTIMMNYRQTESAFWMEYLPTVVGVLIPT YPPTTEFWWEPKEPLQIAFWSLAAACLILMVLVVICCMMWRSAKRESDRLYDEAV FGMDNPESERDTVMENNMLQAQPLRGQSGNIYEYRDSPSKYSKMPPDGGGGSIRS PSSLGMTQVTGKSGLTGSQSSLRSAISMKESSSISPPPLSKESNFDRNIKYAGSDTVP SDTIQRNGKFARSDSKNLLANQMSQAELARSQRDLGGATSPMSETSTMAGSVLGR SQQITPVPSSRGGGMKSNNASRTNLIEGIPQTAV
Culex quinquefasciatus	MVVVAARSSHHGRCRRSWCCVTAPILLIFCTVVHGQYYTTRDPRWYSREGDYNY HLPNPGDPDYRTYTYNNRRYGYYQPSGYGQQYPGQNLPGQYPQNTGLGEDRFKY DPANPNNVQTPFPGVLGGWREDLQGKQRRDWQQLDRDVHVTTNYGQVQGFKVH LYDSPDPKSFYRPWHSPVDRIMGTCATFLGIPYALPPTFEGRFKPPRQHRGWQLLQ AVDYGPACPQPVQYTGATKGIRDMDEDCLYLNVFSPNTQSGIAQKYPVMIYIHGG EFVRGASNTFPGQMLAAFYEVVVVTFNYRLGALGFLSTGDENSPGNYGILDQIMA VRWVYDNIEAFNGDRNSITLFGPGAGGASAGLLMVAPQTKDIITRVIAQSGSAVAD WALIVDKYRAQNTSRVFGQMVGCSIETSWKLVNCMRQGRSFYELGNAEFAPHVG LFPWGPVLDINFTFPGDEWYEGWRERDWHFLRETPEELIRRGHYNRGLHYMSGVT LQEAAFVISQNETLAPYFEVDQKFFEQKVWELVYRYNYTINLNGTYEAIKYMYTY WPDPKNTTFIREQYINLLSDFLYRAPADKMTKLLVERKVPVYSYVMNTTIEGLKLP EWRKVPHDIEHYLLTGAPFMDVEFFPKKARLDRLMWTDNDRNMSHFFMKTYSDF

	ARYGNPTPQQVLGLHFEKAINGELRYLNLNTTYNSSIMMNFRQTESAFWTQYIPTV VGVLVPTYPPTTEFWWEPTEPLQIAFWSMSAACLFLIVLVVICCILWRNAKRQSDR FYDEGVFVVDNPESERGDVGIDNNVLQAQPLRQNIYEYRDSPSKYGKHPPPTDTGS IRSPSSLGTMHTTGKSISGSQSSLRSAISMKEVSVSPAPRSETNFDRNIRYAERGVDT EVQKNGGGKYQRSDSKGTLNDKAREFATSTPIEPVSRSQQPTPQPTTGRGSRAGFD RSNNASRTNLIEGIPQTEV
Tribolium castaneum	MAVITHLKFKLFIFLCSELTCVVLGQNRQYRPNVYGSYDDRNQFSQYNNNPYQIPN PGDRDYRTYVYKGRRYGQQNNFYNNRWRPGDPRIQGQDEHFFYDQQGNGEPILP GVLGGWRTDLQGKFRSQTKDKKRDIFVTTTHGQVQGFLVYLFDNPDPESLYRPGS EFIEREYGMTEVYLGIPYAEPPIVEARFKPPRWHKGWQLLQAVDFGPACPQPARYV GATKGIRDMDEDCLYLNVYKPPTSEEGVGKRFPIMVYIHGGDFIRGASNTFPGHIM ATFYQVIVVTINYRLGALGFISTGDVNSPGNYGILDQAMALKWIYENAESFNGDRY SITLFGPGAGAASAGLLMVAPQTKDMISKVIAQSGSALADWTLIVDKYRAQNTSR VFGKLIGCSIESSWKLMNCLKQGRSFYEIGNSEFPPEVGLFPWAPVMEMNVSMPFY EGWQESDWHFLKETPEDLIRKKHFNPYLKYMSGVTLQEAATFITSNKSLEPNFIIDQ EFFDQKVKELVLRYNYTLNPNGTYEAIKYMYTYWPDPNNKTHIRDKYIQMLSDFL YTAPNDKIVKLLVEQGVEVYMYVLNTTIESFKLEEWRKVPHDIEHYLLCGAPFMD TEFFPEGFSRTQWTNNDRNMSHFFMQAYTNFAKYGNPTYTQILGIHFEVARHGEL KYLNLNTTYNSSIRWNFRQTESAFWSQYLPTVVGHLVPTYPPTTEYWWEPRAPLQI AFWSMSAACLLLVVLLVVCCMLWRNAKRQSDRFIYDPDPDEGVDNTRLTSAYDY AEKLRTPISVPPRTNSASTFRSESAVSLKEAQGFVSSSPSEDYPRKGTPVLSHRHKSK AELAQGVPQTDV
Pediculus humanus	MAIVRDLRQSVQNSLPKPQVLKSERWRGLLVLVGVICVLPMGIRAQWDQYTTRDP RFYSRDGVYNYNPPNPGDPEYRTYTYNDRRYGYYQPAGYGHLYPGQKLPGQYPN PHDTPLGDDRFKFDPSDPNTFQTPFPGVLGGWREDLQGKLRRDSLQLDRDVFVTT NYGQVQGFKVYLYDNPDPLSFYRPWHSTVDRIMGQCSVFLGIPYALPPTLEGRFKP PRPHRGWQLMQAVDYGPACPQPVRYTGATKGIRDMDEDCLYLNVFSPKTGSGVP QKYPVMVYIHGGEFMKGASNLFPAHVLASFYDVVVVSINYRLGALGFLSTGDENS PGNYGILDQAMAIRWVYDNIEAFNGDRHSITLFGPGAGAASAGLLMVAPQTRNIV TRVIAQSGSATADWALIVDKYRAQNTSIVYGKLLGCYIESSYRLVECLRLGRSFYEI ANAEFAPDVGFFPWGPVLDMNFTYPGDNWYEGWREKDWQFLKETPEQLIKKGM FNKGLHYMTGVTTQEAAYAIYQNESLAPYFEIDEEFFDQKVRELTLRYNYTLNLN GTYEAIRYMYTHWPDPKNKTFIREEYINMLSDFIYRAPSDQMTKLLVEQNVPVYM YVLNTTVEGLRLQDWRAVPHNAEHYFLTGAPFMDVEFFPKRDRLDRNMWTDND RNMSHFFMKTFSDFARFGNPTPQQVLGLHFDKAINGELRYLNLNTTFNSTIRMNYR QKQCAFWTQYIPTVVGVLLPTYPPSTEQSDRFYDGDVFMVDNTGADGDGGGIENI SQTHPLRSRENIYEYRDSPTKKSGPPTDISSIRSPSSLATTQISAKTGSQSSLKSAMSL KEGAQNGKDSTYDRNLNNKGKGSDSRNLLADFEKSREAMTSSPVDGVKTAGTPIP ATRKVVKTSKTQLVEGVPQTEV

Figure 3. Gliotactin Homologue Alignment

```
CLUSTAL 2.1 Multiple Sequence Alignments


Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: Dmel         956 aa
Sequence 2: Dana         951 aa
Sequence 3: Dgri         964 aa
Sequence 4: Dpso         949 aa
Sequence 5: Dvir         949 aa
Sequence 6: Dwil         962 aa
Sequence 7: Ccap        1018 aa
Sequence 8: Llon         862 aa
Sequence 9: Ppap         904 aa
Sequence 10: Aaeg         993 aa
Sequence 11: Agam         973 aa
Sequence 12: Cqui         958 aa
Sequence 13: Tcas         846 aa
Sequence 14: Phum         904 aa
```

---

clustalw.aln

```
CLUSTAL 2.1 multiple sequence alignment


Ppap            ------------------------MAIVRDLRQSVQNSLPKPQVLKSERWRGLLVLVGVI
Phum            ------------------------MAIVRDLRQSVQNSLPKPQVLKSERWRGLLVLVGVI
Llon            -------------------------MAQDLRQSAQNSKTKPQKRISVRWRGLAVIVGVM
Aaeg            MALNTNCGAAKRYQQAKHSAKTTKSSDGGGGCSVRKKCVHLSSSTTGSEWIHAKINVLFL
Cqui            -------------------------------------MVVVAARSSHHGRCRRSWCCVTAP
Agam            ----------------------------------------MTLSREHHHHASLLLLVVIV
Dgri            --------------------------------MSEQQLQLATMSQKFKYRDKFECLSGILM
Dvir            ----------------------------------------MAQKFKYRDKLKCLLAMLM
Dana            ----------------------------------------MMHKLKYRDKLKWLQALLI
Dpso            ----------------------------------------MMHKLKYRDKSKWLTAILF
Dmel            ----------------------------------------MMHKLKYRDKLKWLLALLV
Dwil            ----------------------------------------MLHKLKYRDKLKCSYAICL
Ccap            -------------------------MVHRKVVQLINTRPLRMNTFQYHDKVKSAVMLVI
Tcas            -------------------------------------MAVITHLKFKLFIFLCSELTCV


Ppap            CVLPMGIRAQWDQYTTRDPRFYSRDGVYNYNPPNPGDPEYRTYTYNDRRYGYYQPAGYGH
Phum            CVLPMGIRAQWDQYTTRDPRFYSRDGVYNYNPPNPGDPEYRTYTYNDRRYGYYQPAGYGH
Llon            CVLPVGIRAQWDQYTTRDPRFYSRDGVNNYNPPNPGDPEYRTYTYNDRRYGYYQPAGYGH
Aaeg            CFAVFISLVDGQYHTTRDPRWYSREGDFSYHLPNPGDPDYRTYTYNNRRYGYYQPNGYGQ
Cqui            ILLIFCTVVHGQYYTTRDPRWYSREGDYNYHLPNPGDPDYRTYTYNNRRYGYYQPSGYGQ
Agam            GVASLVGVCDGQYYTTRDPRWYSREGDFNYKIPNPGDPEYRTYTFNNRRYGYYQPNGYGQ
Dgri            VLLVLQPLLPHIDAQTRDPRFYSRVGV-DYQWPNPGDPEYRTYTFNDRRYGHYQPNGYGA
Dvir            ALMLPQ----SIDAQTRDPRFYSRVGV-DYQWPNPGDPDYRTYTFNDRRYGHYQPNGYGA
Dana            CVCLSH----FAEAQTRDPRFYSRP---GVDWPNPGDPDYRTYTFNDRRYGHYQPNGFGA
Dpso            LVVASLG--PLVGGQTRDPRFYSRPGV-DYHWPNPGDPDYRTYTFNDRRYGHYQPNGFGA
Dmel            LIGTCF---IQTRGQTRDPRFYSRPGV-DYHWPNPGDPDYRTYTFNDRRYGHYQPNGYGA
Dwil            ILILSFLATLTDGQTRDPRFYSRVGVN-DYHWPNPGDPDYRTYTFNDRRYGHYQPNGFGA
Ccap            ASYLVLSNLICVKAQTRDPRFFSRPGVNDYNWPNPGDPDYRTYIFNDRRYGHYLPNGYGT
Tcas            VLGQNRQYRPNVYGSYDDRNQFSQYNNNPYQIPNPGDRDYRTYVYKGRRYGQQNNFYNNR
                                          . ***** :**** ::.****       .

Ppap            LYPGQKLPGQYPNPHDTPLGDDRFKFDPSDPNTFQTPFPGVLGGWREDLQGKLRRDSLQL
Phum            LYPGQKLPGQYPNPHDTPLGDDRFKFDPSDPNTFQTPFPGVLGGWREDLQGKLRRDSLQL
Llon            LYPGQKLPGQYPNPHDTPLGDDRFKFDPSDPNTFQTPFPGVLGGWREDLQGKLRRDSLTL
Aaeg            QYPGQNLPGQYP--QNTGLGEDRFKYDPTNPNNVQTPFPGVLGGWREDLQGKLRRDSIQL
Cqui            QYPGQNLPGQYP--QNTGLGEDRFKYDPANPNNVQTPFPGVLGGWREDLQGKQRRDWQQL
Agam            QYPGQSLPGQYP--SNTGLGDDRFKYDPTNPTLAQTPFPGVLGGWREDLQGKLRRDSMQL
Dgri            NYPGRNPPGQYP---QGMPNEDRFRFDPNDP-NARTQFPGVLAGWREDLQGKQRRDSLTL
Dvir            NYPGRNPPGQYP---QGMPNEDRFRFDPNDP-NARTQFPGVLAGWREDLQGKQRRDSLTL
Dana            NYPGRNSPGQYP---QGMPNEDRFRFDPNDP-NARTQFPGVLAGWREDLQGKQRRDSLTL
Dpso            NYPGRNPPGQYP---QGMPNEDRFRFDPNDP-NARTQFPGVLAGWREDLQGKQRRDSLTL
```

```
Dmel            NYPGRNPPGQYP---QGMPNEDRFRFDPNDP-NARTQFPGVLAGWREDLQGKQRRDSLTL
Dwil            NFPGRNPPGQYP---QGMPNEDRFRFDPNDP-NARTQFPGVLAGWREDLQGKQRRDSLTL
Ccap            NYPGRNSPGQYP---QGMPNEDRFRFDPNNPNAAHTPFPGILAGWREDLQGKQRRDSLTL
Tcas            WRPGDPRIQGQD-----------EHFFYDQQGNGEPILPGVLGGWRTDLQGKFRSQTKDK
                  **                    ::   :    .. :**:*.*** ***** * :

Ppap            DRDVFVTTNYGQVQGFKVYLYDNPDPLSFYRPWHSTVDRIMGQCSVFLGIPYALPPTLEG
Phum            DRDVFVTTNYGQVQGFKVYLYDNPDPLSFYRPWHSTVDRIMGQCSVFLGIPYALPPTLEG
Llon            DRDVFVTTNYGQVQGFKVYLYDNPDPLSFYRPWHSTVDRIMGQCSVFLGIPYALPPTHEG
Aaeg            DRDVFVTTNYGQVQGFKVHLYDNPDPKSFFRPWHSPVDRVMGTCSTFLGIPYALPPTFEG
Cqui            DRDVHVTTNYGQVQGFKVHLYDSPDPKSFYRPWHSPVDRIMGTCATFLGIPYALPPTFEG
Agam            ERDVFVTTNYGQVQGFMVYLYDNPDPKNLYRPWHSNVDRIMGKCSTFLGIPYALPPTHEG
Dgri            ERDVFVTTNYGQVQGFKVYMYDNPDPKSFYRPYHSTVDRVMGECSTFLGIPYALPPTFEG
Dvir            ERDVFVTTNYGQVQGFKVYMYDNPDPKSFYRPYHSTVDRVMGECSVFLGIPYALPPTFEG
Dana            ERDVFVTTNYGQVQGFKVYMYDNPDPKSFYRPYHSTVDRVMGECSAFLGIPYALPPTFEG
Dpso            DRDVFVTTNYGQVQGFKVYMYDNPDPKSFYRPYHSTVDRVMGECSAFLGIPYALPPTFEG
Dmel            ERDVFVTTNYGQVQGFKVYMYDNPDPKSFYRPYHSTVDRVMGECSVFLGIPYALPPTFEG
Dwil            ERDVFVTTNYGQVQGFKVYMYDNPDPKSFYRPYHSTVDRVMGECSVFLGIPYALPPTFEG
Ccap            DRDVFVTTNYGQVQGFKVYMYDNPDPKSFYRPYHSTVDRVMGECSVFLGIPYALPPTFEG
Tcas            KRDIFVTTTHGQVQGFLVYLFDNPDPESLYRPGSEFIEREYGMTEVYLGIPYAEPPIVEA
                .**:.***.:****** *:::*.*** .::**  . ::*  *   .:****** **  *.

Ppap            RFKPPRPHRGWQLMQAVDYGPACPQPVRYTGATKGIRDMDEDCLYLNVFSPKTG-SGVPQ
Phum            RFKPPRPHRGWQLMQAVDYGPACPQPVRYTGATKGIRDMDEDCLYLNVFSPKTG-SGVPQ
Llon            RFKPPRPHRGWQLMQAVDFGPACPQPVRYTGATKGIRDMDEDCLYLNVFSPKTG-SGVPQ
Aaeg            RFKPPRPHRGWQLLQAVDYGPACPQPVQYTGATKGIRDMDEDCLYLNVFSPNTQ-SGVAQ
Cqui            RFKPPRQHRGWQLLQAVDYGPACPQPVQYTGATKGIRDMDEDCLYLNVFSPNTQ-SGIAQ
Agam            RFKPPRQHRGWQLVQAVDYGPACPQPVQYTGATKGIRDMDEDCLYLNIFSPNTQ-AGVAQ
Dgri            RFKPPRLHRGWQLLQAVDFGPACPQPVRYTGATKGVMDMDEDCLYLNVYSPKTG-AGVAQ
Dvir            RFKPPRLHRGWQLLQAVDFGPACPQPVRYTGATKGVMDMDEDCLYLNVYSPKTG-AGVAQ
Dana            RFKPPRIHRGWQLLQAVDFGPACPQPVRYTGATKGVMDMDEDCLYLNVYSPKTG-AGVAQ
Dpso            RFKPPRAHRGWQLLQAVDFGPACPQPVRYTGATKGVMDMDEDCLYLNVYSPKTG-AGVAQ
Dmel            RFKPPRVHRGWQLLQAVDFGPACPQPVRYTGATKGIMDMDEDCLYLNVYSPKTG-AGVAQ
Dwil            RFKPPRPHRGWQLLQAVDFGPACPQPVRYTGATKGVMDMDEDCLYLNVYSPKTG-AGVAQ
Ccap            RFKPPRLHRGWQLLQAVDYGPACPQPVRYTGATKGIMDMDEDCLYLNIFSPKTG-AGVAQ
Tcas            RFKPPRWHKGWQLLQAVDFGPACPQPARYVGATKGIRDMDEDCLYLNVYKPPTSEEGVGK
                ****** *:****:****:*******.:*.*****: **********::.* *   *: :

Ppap            KYPVMVYIHGGEFMKGASNLFPAHVLASFYDVVVVSINYRLGALGFLSTGDENSPGNYGI
Phum            KYPVMVYIHGGEFMKGASNLFPAHVLASFYDVVVVSINYRLGALGFLSTGDENSPGNYGI
Llon            KYPVMVYIHGGEFTRGASNLFPAHVLASFYEVVVVSINYRL-------------------
Aaeg            RFPVMIYIHGGEFVRGASNVFPGHMLAAFYEVVVVTFNYRLGALGFLSTGDENSPGNYGI
Cqui            KYPVMIYIHGGEFVRGASNTFPGQMLAAFYEVVVVTFNYRLGALGFLSTGDENSPGNYGI
Agam            RYPVMIYIHGGEFVRGASNTFPGHMLAAFYDVVVVTFNYRLGALGFLSTGDENSPGNYGI
Dgri            KYPVMVYLHGGEFVRGASNLFQGHILASFYDVVVVTLNYRLGALGFLSTADENSPGNYGI
Dvir            KYPVMVYIHGGEFVHGASNLFPGHVLASFYDVVVVTLNYRLGALGFLSTGDENSPGNYGI
Dana            KYPVMVYIHGGEFIRGASNLFQGHILASFYDVVVVTLNYRLGALGFLSTGDENSPGNYGI
Dpso            KYPVMVYIHGGEFIRGASNLFQGHILASFYDVVVVTLNYRLGALGFLSTGDENSPGNYGI
Dmel            KYPVMVYIHGGEFIRGASNLFQGHILASFYDVVVVTLNYRLGALGFLSTGDENSPGNYGI
Dwil            KYPVMVYIHGGEFIRGASNLFQGHMLASFYDVVVVTLNYRLGALGFLSTGDENSPGNYGI
Ccap            KYPVMVYIHGGEFIRGASNLFQGHILASFYDVVVVTINYRLGALGFLSTGDDNSPGNYGI
Tcas            RFPIMVYIHGGDFIRGASNTFPGHIMATFYQVIVVTINYRLGALGFISTGDVNSPGNYGI
                ::*:*:*:***:* :**** * .:::*:**:*:**::****

Ppap            LDQAMAIRWVYDNIEAFNGDRHSITLFGPGAGAASAGLLMVAPQTRNIVTRVIAQSGSAT
Phum            LDQAMAIRWVYDNIEAFNGDRHSITLFGPGAGAASAGLLMVAPQTRNIVTRVIAQSGSAT
Llon            ------------------------------------------------------------AT
Aaeg            LDQIMAVRWVYDNIEAFNGDRNSITLFGPGAGAASAGLLMVAPQTKDIVTRVIAQSGAAV
Cqui            LDQIMAVRWVYDNIEAFNGDRNSITLFGPGAGGASAGLLMVAPQTKDIITRVIAQSGSAV
Agam            LDQIMAVRWVYENIESFNGDRNSITLFGPGAGGASAGLLMVAPQTKDIVTRVIGQSGSAL
Dgri            LDQAMALKWVYDNIEFFNGDRESITLFGPGAGAASAGLLMVAPQTRNIVKRVIAQSGSAL
Dvir            LDQSMALKWVHDNIEFFNGDRDSITLFGPGAGGASAGLLMVAPQTRNIVKRVIAQSGSAL
Dana            LDQAMALRWIHDNIEFFNGDRDSITLFGPGAGGASAGLLMVAPQTRNIVKRVIAQSGSAL
Dpso            LDQAMALRWIHDNIEFFNGDRDSITLFGPGAGGASAGLLMVAPQTRHMVKRVIAQSGSAL
Dmel            LDQAMALRWVYDNIEFFNGDRNSITLFGPGAGGASAGLLMVAPQTRNIVRRVIAQSGSAL
Dwil            LDQALALKWVYDNIEFFNGDRESITLFGPGAGGASAGLLMVAPQTRNIVKRVIAQSGSAL
Ccap            LDQVMALKWIYDNVEFFNGDRESITLFGPGAGAASAGLLMVAPQTKNIVKRVIAQSGAAM
Tcas            LDQAMALKWIYENAESFNGDRYSITLFGPGAGAASAGLLMVAPQTKDMISKVIAQSGSAL
                                                                          *

Ppap            ADWALIVDKYRAQNTSIVYGKLLGCYIESSYRLVECLRLGRSFYEIANAEFAPDVGFFPW
```

```
Phum            ADWALIVDKYRAQNTSIVYGKLLGCYIESSYRLVECLRLGRSFYEIANAEFAPDVGFFPW
Llon            AEWALTVDKYRAQNTSIVYGKLLGCYIESSYRLVECLRLGRSFYEIANAEFAPDVGFFPW
Aaeg            ADWALIVDKYRAQNTSRVFGQMVGCSIETSWKLVNCMRQGRSFFELGNAEFAPHVGLFPW
Cqui            ADWALIVDKYRAQNTSRVFGQMVGCSIETSWKLVNCMRQGRSFYELGNAEFAPHVGLFPW
Agam            ADWALIVDKWRAQNTSRVFAQNVGCSIETSWKMVNCMRNGRSSYELGNAEFPPHVGLFPW
Dgri            ADWAFIQDKYRAQNTSRVLGQVLGCSMDSSWKLVNCLRTGRSFYELGNAEFSPQVGSFPW
Dvir            ADWALIQDKYRAQNTSRVLGQVLGCSIDSSWKLVNCLRTGRSFYELGNAEFPPQVGSFPW
Dana            ADWALIQDKYRAQNTSRVLGQLLGCSIESSWKLVNCLRTGRSFYELGNAEFAPQVGSFPW
Dpso            ADWALIQDKYRAQNTSRVLGQLLGCSIESSWKLVNCLRTGRSFYELGNAEFSPQVGSFPW
Dmel            ADWALIQDKYRAQNTSRVLGQLLGCSIESSWKLVNCLRTGRSFYELGNAEFSPQVGSFPW
Dwil            ADWALIQDKYRAQNTSRVLGQLLGCSIESSWKLVNCLRTGRSFYELGNAEFSPQVGSFPW
Ccap            ADWALIQDKYRAQNTSRVLGQLLGCSIESSWKLVNCLRTGRSFYELGNAEFPPQVGTFPW
Tcas            ADWTLIVDKYRAQNTSRVFGKLIGCSIESSWKLMNCLKQGRSFYEIGNSEFPPEVGLFPW
                *:*::  **:****** * .: :** :::*::::::*:: *** :*:.*:**.*.** ***

Ppap            GPVLDMNFTYPGDNWYEGWREKDWQFLKETPEQLIKKGMFNKGLHYMTGVTTQEAAYAIY
Phum            GPVLDMNFTYPGDNWYEGWREKDWQFLKETPEQLIKKGMFNKGLHYMTGVTTQEAAYAIY
Llon            GPVLDMNLTYPGDNWYEGWREKDWQFLKETPEQMIKKGLFNKGLHYMTGVTTQEAAYVLY
Aaeg            GPVLDINFTFPGDEWYEGWRERDWHFLAETPEELIRKGHYNRGLHYMSGVTLQEAAFVIS
Cqui            GPVLDINFTFPGDEWYEGWRERDWHFLRETPEELIRRGHYNRGLHYMSGVTLQEAAFVIS
Agam            GPVLENNFTFPGDSWYEGWSERDWHFLAETPESLIRRRHFNRGLHYMSGVTLQEAAFVIT
Dgri            GPVLDHNFTLPGDDWYEGWREKDWRFLTQTPETLIRAGRFNRNIQYMTGVTTQEAAFFVA
Dvir            GPVLDHNFTLPGDDWYEGWREKDWRFLTQTPETLIRSGRFNRNIQYMTGVTTQEAAFFVA
Dana            GPVLDHNFTVPGDDWYEGWREKDWRFLTQTPETLIRAGRFNRNIQYMTGVTTQEAAFLVA
Dpso            GPVLDHNFTLPGDDWYEGWREKDWRFLTQTPETLIRAGKFNRNIQYMTGVTTQEAAFLVA
Dmel            GPVLDHNFTLPGDDWYEGWREKDWRFLTQTPETLIRAGKFNRNIQYMTGVTTQEAAFFVA
Dwil            APVLDHNFTLPGDDWYEGWRQKDWRFLTQTPETLIRAGRFNRNIQYMTGVTTQEAAFFIA
Ccap            GPVLDHNFTVPGDDWYEGWREKDWRFLTQTPEYLIRQGHFNRGLQYMTGVTTQEASFFVA
Tcas            APVMEMNVSMP---FYEGWQESDWHFLKETPEDLIRKKHFNPYLKYMSGVTLQEAATFIT
                .**:: *.: *   :**** : **:** :*** :*:   :*  ::**:*** ***:  :

Ppap            QNESLAPYFEIDEEFFDQKVRELTLRYNYTLNLNGTYEAIRYMYTHWPDPKNKTFIREEY
Phum            QNESLAPYFEIDEEFFDQKVRELTLRYNYTLNLNGTYEAIRYMYTHWPDPKNKTFIREEY
Llon            QNESLAPYFDIDEDFVDQKIRELVLRYNYTLNLNGTYEAIKYMYTHWPDPSNKTYIREEY
Aaeg            QNESLAPYFEVDQKFFEQKVWELVYRYNYTLNLNGTYEAIKYMYTHWPDPRNTTYIREQY
Cqui            QNETLAPYFEVDQKFFEQKVWELVYRYNYTINLNGTYEAIKYMYTYWPDPKNTTFIREQY
Agam            GNKSLAPYFEVDERFFDQKVRELVLRYNYTLNLNGTYEAIKYMYTYWPDPKNTTFIREQY
Dgri            QNESLAPYYELDGRFFEQKVREHVFRYNYTLNPNGVYEAIKYMYTFWPDPNNNTIIRDQY
Dvir            QNESLAPYYELDNRFFEQKVREHVFRYNYTLNPNGVYEAIKYMYTYWPDPNNNTIIRDQY
Dana            QNESLSPYYELDGRFFEQKIREHVFRYNYTLNPNGVYEAIKYMYTYWPDPNNNTIIRDQY
Dpso            QNESLGPYYELDGRFFEQKIREHVFRYNYTLNPNGVYEAIKYMYTYWPDPNNNTIIRDQY
Dmel            QNESLSPYYELDGRFFDQKIREHVFRYNYTLNPNGVYEAIKYIYTFWPDPNNNTIIRDQY
Dwil            QNESLAPYYEIDGRFFEQKVREHVFRYNYTLNPNGVYEAIKYMYTFWPDPNNNTIMRDEY
Ccap            QNESLAPYYEIDHTFFDAKIREHVFRYNYTLNPNGVYEAIKYMYTYWPDPNNSSVIRDQY
Tcas            SNKSLEPNFIIDQEFFDQKVKELVLRYNYTLNPNGTYEAIKYMYTYWPDPNNKTHIRDKY
                 *::* * : :*  *.: *: * . *****:* **.****:*:**.**** *.: :*::*

Ppap            INMLSDFIYRAPSDQMTKLLVEQNVPVYMYVLNTTVEGLRLQDWRAVPHNAEHYFLTGAP
Phum            INMLSDFIYRAPSDQMTKLLVEQNVPVYMYVLNTTVEGLRLQDWRAVPHNAEHYFLTGAP
Llon            INLLSDFIYRAPSDQMTKLLVEQNVPVYMYVLNTTVEGMRLQNWRAVPHNIEHYFLTGAP
Aaeg            INLLSDFLYRAPADKMTKLLVERRVPVYSYVMNTTIEGLKLPEWRKVPHDIEHYLLTGAP
Cqui            INLLSDFLYRAPADKMTKLLVERKVPVYSYVMNTTIEGLKLPEWRKVPHDIEHYLLTGAP
Agam            INLLSDFLYRAPTDKLIKLLVEQNVPVYSYVMNTTIEGLKLAEWCKVPHDIEHYLLAGAP
Dgri            INMLSDLYYRAPVDQMVKLLLEQKVPVFMYVLNTTVEALNLPQWRKYPHDTEHYFLTGAP
Dvir            INMLSDLYYRAPVDQMVKLLLEQKIPVYMYVLNTTVEALNLPQWRKYPHDTEHYFLTGAP
Dana            INLLSDLYYRAPVDQMVKLLLEQKVPVYMYVLNTTVEALNLPQWRKYPHDIERYFLTGAP
Dpso            INMLSDLYYRAPVDQMVKLLLEQKVPVFMYVLNTTVEALNLPQWRKYPHDTERYFLTGAP
Dmel            INMLSDLYYRAPVDQMVKLMLEQKVPVYMYVLNTTVEALNLPQWRKYPHDIERYFLTGAP
Dwil            INMLSDLYYRAPVDQMVKLMLEQKVPVYMYVLNTTVEALNLPQWRKYPHDTEHYFLTGAP
Ccap            INLLSDLYYRAPVDKMVKLLLEQRIPVYMYVLNTTVEALNYPQWRKYPHNIEHYFLTGAP
Tcas            IQMLSDFLYTAPNDKIVKLLVEQGVEVYMYVLNTTIESFKLEEWRKVPHDIEHYLLCGAP
                *::***: * ** *:: **::*: : *: **:***:*.:.  :*   **: *:*:* ***

Ppap            FMDVEFFPKRDRLDRNMWTDNDRNMSHFFMKTFSDFARFGNPTPQQVLGLHFDKAINGEL
Phum            FMDVEFFPKRDRLDRNMWTDNDRNMSHFFMKTFSDFARFGNPTPQQVLGLHFDKAINGEL
Llon            FMDVEFFPKRDRLERNMWTDNDRNMSHFFMKTFSDFARFGNPTPQQVLGLHFDKAINGEL
Aaeg            FMDVEFFPRKARLDRLMWTDNDRNMSHFFMKTFSDFARYGNPTPQQVLGLFFEPARNGEL
Cqui            FMDVEFFPKKARLDRLMWTDNDRNMSHFFMKTYSDFARYGNPTPQQVLGLHFEKAINGEL
Agam            FMDVEFFPRRERFDRLMWTDNDRNMSHFFMKTYSDFARHGNPTPTQVLGLYFEKARNGEL
Dgri            FMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYSNFARYGNPTPQQVLGMHFQRAYQGEI
Dvir            FMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYSNFARYGNPTPQQVLGMHFQRAYQGEI
```

```
Dana            FMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYSNFARYGNPTPQQVLGMHFQRAYQGEI
Dpso            FMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYSNFARYGNPTPQQVLGMHFQRAYQGEI
Dmel            FMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYTNFARYGNPTPQQVLGMHFQRAYQGEI
Dwil            FMDTEFFPKKEHLQRNMWTDNDRNMSHFFMQTYSNFARYGNPTPQQVLGMHFQRAYQGEI
Ccap            FMDIEFFPKKEHLQRNMWTDNDRNMSHFFMQTYSNFARYGNPTPQQVLGLHFQRAYQGEL
Tcas            FMDTEFFPEG--FSRTQWTNNDRNMSHFFMQAYTNFAKYGNPTYTQILGIHFEVARHGEL
                *** ****.   :.*  **:**********:::::**:.****  *:**:.*: * :**:

Ppap            RYLNLNTTFNSTIRMNYRQKQCAFWTQYIPTVVGVLLPTYPPSTE---------------
Phum            RYLNLNTTFNSTIRMNYRQKQCAFWTQYIPTVVGVLLPTYPPSTE---------------
Llon            RYLNLNTTFNSTIRMNYRQKQCAFWTQYLPTVVGVLVPTYPPSTEYWWEPQQPLQIAFWS
Aaeg            KYLNLNTTYNSSIRMNYRQTESAFWTQYIPTVVGVMVPTYPPTTEFWWEPKEPLQIAFWS
Cqui            RYLNLNTTYNSSIMMNFRQTESAFWTQYIPTVVGVLVPTYPPTTEFWWEPTEPLQIAFWS
Agam            KYLNLNTTYNSTIMMNYRQTESAFWMEYLPTVVGVLIPTYPPTTEFWWEPKEPLQIAFWS
Dgri            RYLNINTTYNSSILLNYRQTESAFWTQYLPTVIGVLVPTYPPTTEYWWEPKEPLQIAFWS
Dvir            RYLNINTTYNSSILLNYRQTECAFWTQYLPTVIGVLVPTYPPTTEYWWEPKEPLQIAFWS
Dana            RYLNVNTTFNSSILLNYRQTECAFWTQYLPTVIGVLVPTYPPTTEYWWEPKEPLQIAFWS
Dpso            RYLNVNTTFNSSILLNYRQTECAFWTQYLPTVIGILVPTYPPTTEYWWEPKEPLQIAFWS
Dmel            RYLNINTTYNSSILLNYRQTECAFWTQYLPTVIGVLVPTYPPTTEYWWEPKEPLQIAFWS
Dwil            RYLNINTTYNSSILLNYRQTECAFWTQYLPTVIGVLVPTYPPTTEYWWEPKEPLQIAFWS
Ccap            RYLNINTTFNSSILLNYRQTECAFWTQYLPTVIGVLVPTYPPTTEFWWEPKEPLQIAFWT
Tcas            KYLNLNTTYNSSIRWNFRQTESAFWSQYLPTVVGHLVPTYPPTTEYWWEPRAPLQIAFWS
                :***:***:**:*  *:**.:.*** :*:***:* ::*****:**

Ppap            ------------------------QSDRFYDGDVFMVDNTGADG-DGGGIENISQTHPLR
Phum            ------------------------QSDRFYDGDVFMVDNTGADG-DGGGIENISQTHPLR
Llon            ISAVCLFLVILLVVCCILWRNAKRQSDRFYDGDVFMVDNTGADG-DAG-IENISQTHPLR
Aaeg            MSAACLLLIVMVVICCILWRNAKRQSDRFYDDGVIVVDNPESER-DVGIDNNVLQAQPLR
Cqui            MSAACLFLIVLVVICCILWRNAKRQSDRFYDEGVFVVDNPESERGDVGIDNNVLQAQPLR
Agam            LAAACLILMVLVVICCMMWRSAKRESDRLYDEAVFGMDNPESER-DTVMENNMLQAQPLR
Dgri            MSVACFFLIVLVVICCIMWRNAKRQSDRFYDEDVFINGEGMEPEPDPRGVDNAHMVTNHH
Dvir            MSVACFFLIVLVVICCIMWRNAKRQSDRFYDEDVFINGEGMEPEPDARGVDNAHMVTNHH
Dana            MSVTCFFLIVLVVICCIMWRNAKRQSDRFYDEDVFINGEGMEPEPDARGVDNAHMVTNHH
Dpso            MSVTCFFLIVLVVICCIMWRNAKRQSDRFYDEDVFINGEGLEPEPDPRGVDNAHMVTNHH
Dmel            MSVACFFLIVLVVICCIMWRNAKRQSDRFYDEDVFINGEGLEPEQDTRGVDNAHMVTNHH
Dwil            MSVTCFFLIVLVVICCIMWRNAKRQSDRFYDEDVFINGDGIEPETDPRGVDNAHMVTNHH
Ccap            MSVACFFLIVLVVICCIMWRNAKRRSDFYDDDVFINPDGAELPDEVHTGVDNTHMVANPH
Tcas            MSAACLLLVVLLVVCCMLWRNAKRQSDRFIYDPDP------------------------
                                  .**

Ppap            ---SRENIYEYRDSPTKKSGPPTDISSIRSPSSLATTQISAKTGSQSSLKSAMSLKEG--
Phum            ---SRENIYEYRDSPTKKSGPPTDISSIRSPSSLATTQISAKTGSQSSLKSAMSLKEG--
Llon            ---SRENIYEFRDTPKK-SGPPTDISSVRSPSSLATNPT-AKSGSQSSLKSAISLKEG--
Aaeg            ---TRDNIYEYRDSPTKYKSPPTDATSIRAPSSLAMTQTTAKS-ISGSQSSLRSAISMKE
Cqui            ---QNIYEYRDSPSKYGKHPPPTDTGSIRSPSSLGTMHTTGKS-ISGSQSSLRSAISMKE
Agam            GQSGNIYEYRDSPSKYSKMPPDGGGGSIRSPSSLGMTQVTGKSGLTGSQSSLRSAISMKE
Dgri            AMRSRDNIYEYRDSPSTKTLASKVHTDTTSLRSPSSLAMTQKSSSQASIKSGISLKETNG
Dvir            ALRSRDNIYEYRDSPSTKTLASKAQTDTTSLRSPSSLAMTQKSGSQASLKSGISLKETNG
Dana            ALRSRDNIYEYRDSPSSKTLASKAQTDTTSLRSPSSLAMTQKSGSQASLKSGISLKETNG
Dpso            AMRSRDNIYEYRDSPSSKTLASKAHTDTTSLRSPSSLATTQKSGSQASLKSGISLKETNG
Dmel            ALRSRDNIYEYRDSPSTKTLASKAHTDTTSLRSPSSLAMTQKSSSQASLKSGISLKETNG
Dwil            ALRSRDNIYEYRDSPSSKTLASKAHTDTTSLRSPSSLAMTQKSGSQASLKSGISLKETNG
Ccap            AMRSRDNIYEYRDSPSSKTLASKIHTDTTSIRSPSSLAMTQKSGSQSSLKSTMSLKETTG
Tcas            ------------DEGVDNTRLTSAYDYAEKLRTPISVPPRTNSASTFRSESAVSLKEAQG
                                          :       ::      .*  *  .

Ppap            --------------------------------AQNGKDSTYDRNLNNKGKGSDSRNLLADF
Phum            --------------------------------AQNGKDSTYDRNLNNKGKGSDSRNLLADF
Llon            --------------------------------AQNGRDGNYDRNLNNKGKGSDGKNLLADF
Aaeg            STNISPIPRS-ESNFD----------RNIKYAERGVDTEINKNGKYPRTDSRG-ALSDTK
Cqui            VS-VSPAPRS-ETNFD----------RNIRYAERGVDTEVQKNGGGKYQRSDSKGTLNDK
Agam            SSSISPPPLSKESNFD----------RNIKYAGSDTVPSDTIQRNGKFARSDSKNLLANQ
Dgri            SLVK--SERAATPRS-----------QVNGTVSSKAPPVEEKRLLQPISSTPVAQLQSE-
Dvir            SLVK--SERAATPRS-----------QTN-GTSAKAAPIEEKRLLQPLSSTPVAQLQAE-
Dana            SLTKSVSARSQANGN-----------GTTAKSAAAAP-VEEKRLLQPLSSTPVPQLQAE-
Dpso            SLVK--SERAATPRS-----------QPNGTTAKAGP-VEEKRLLHPLSSTPVPQLQSE-
Dmel            HLVKQSERAATPRSQ-----------QNGSTAKVASPPVEEKRLLQPLSSTPVTQLQAE-
Dwil            SLVK--SERAATPRS-----------QTNGKSPNPGP-IEEKRLLQPISSTPVPQLQAE-
Ccap            GTLTSCSSSARPSRTNGLHNGSAATIQTPKVFEEKRLITRELTKPEPISSTPLVTTSTA-
Tcas            FVSS-----------------------------------SPSEDYPRKGTPVLSHRHK-
                                                        :
```

```
Ppap            EKSREAMTSSPVDG-------------------------------------VKTAG-TPIP
Phum            EKSREAMTSSPVDG-------------------------------------VKTAG-TPIP
Llon            EKSREAMTSSPVDV-------------------------------------KTSSHGIPIP
Aaeg            SREFSASTPIEPAP-------------------------------------RSQQPSSQPA
Cqui            AREFATSTPIEPVS-------------------------------------RSQQPTPQPT
Agam            MSQAELARSQRDLG-------------------------------------GATSPMSETS
Dgri            --------------PTKRTPTASSVASSS-----------------RSTTPVPSAR---T
Dvir            --------------PTKRMAAAPSVASSS-----------------RSTTPVPSAR---T
Dana            --------------PAKRVPTAASVSGSS-----------------RSTTPVPSARTT-T
Dpso            --------------PTKRAP--ASVSGSS-----------------RSTTPVPSAR---T
Dmel            --------------PAKRVPTAASVSGSS-----------------RSTTPVPSARSTTT
Dwil            --------------PVKRTTPGPAASISG-----------------SSRSNTPVPMARTT
Ccap            --------------KISITSEPISLKPTSNIIATSTSPNLTGTVSSRSTTPVPSARTHTH
Tcas            ------------------------------------------------------------


Ppap            ATR----------------------KVVKTSKTQLVEGVPQTEV
Phum            ATR----------------------KVVKTSKTQLVEGVPQTEV
Llon            ASR----------------------KTVKTSKTQLVEGVPQTEV
Aaeg            SRG--------------ASRAGYDRSNNASRTNLIEGIPQTEV
Cqui            TGR--------------GSRAGFDRSNNASRTNLIEGIPQTEV
Agam            TMAGSVLGRSQQITPVPSSRGGGMKSNNASRTNLIEGIPQTAV
Dgri            TTTTATLTTQPAAQ--------------PRRTQLVEGVPQTSV
Dvir            TTTTATLTAQPAVQ--------------PRRTQLVEGVPQTSV
Dana            TTTTATLSSQPAPQ--------------PRRTQLVEGVPQTSV
Dpso            TTTTATLSSQPAAQ--------------PRRTQLVEGVPQTSV
Dmel            HTTTATLSSQPAAQ--------------PRRTHLVEGVPQTSV
Dwil            TTTTATLTTQPPAAGRS--------VDQPRRTELVEGIPQTSV
Ccap            TTTATVTSAPQVQRPVLPTSSAQAQSRAQARTQVIEGVPQTSV
Tcas            ------------------------------SKAELAQGVPQTDV
                                             ::.: :*:*** *
```

Figure 3. Gliotactin Homologue Alignment. The yellow, highlighted regions indicate the transmembrane domain for each protein sequence.

Figure 4. Gliotactin Phylogenetic Tree

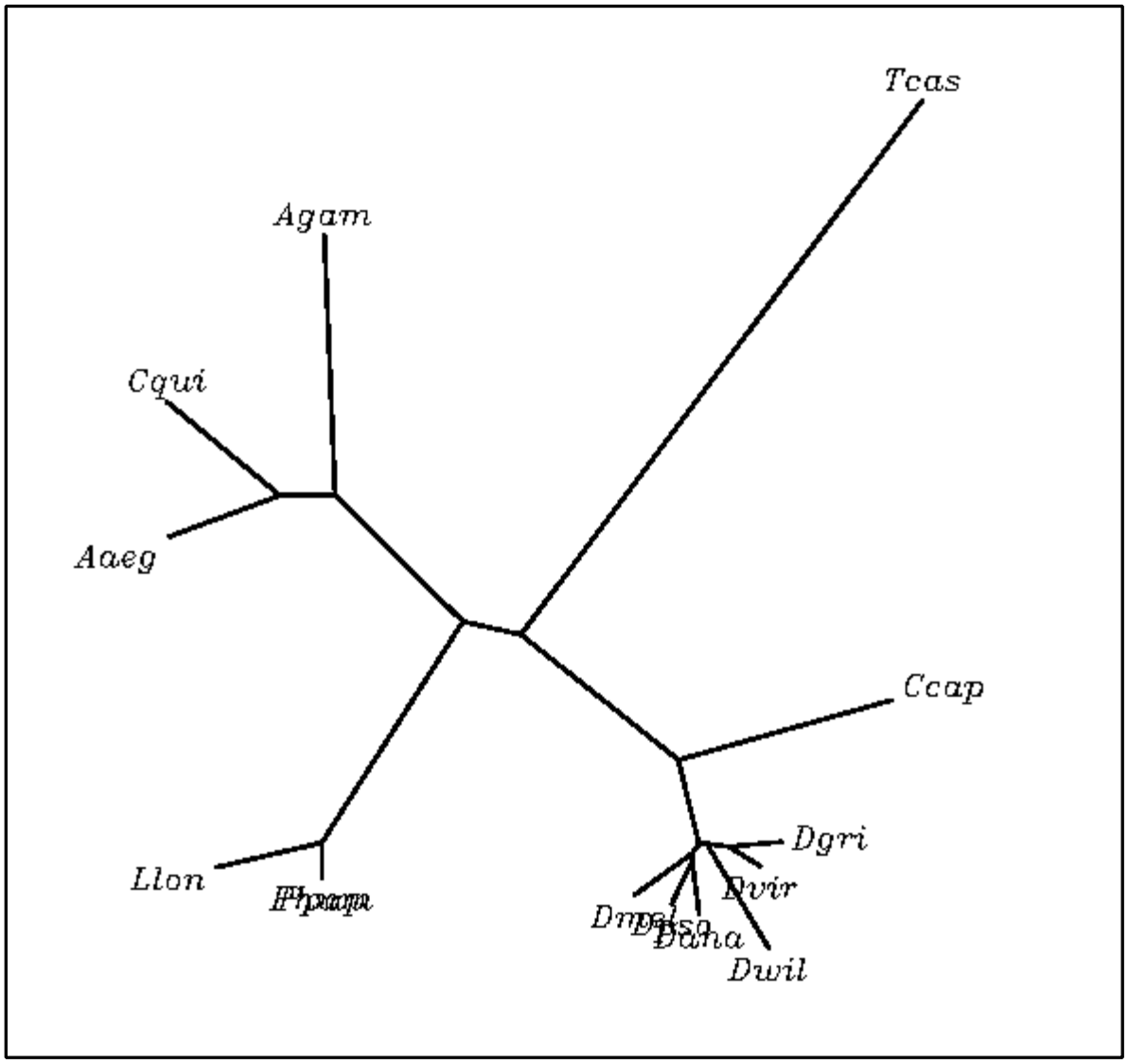


Table 2. Neuroligin 1 Homologues

D. melanogaster	MKFRLAAFWLFLLTVGGNHKLLSHVSPMGVAAEKQPLHKGNELENAMKLREAPK QSRVIGDITTTIQPDSDPGRSIGHQALRRAKAPPPSLELHFRQNLNKLFAGDETTHV GHTNATDNTTAAVLATDEGSSEHEPSTTHHPERRHVVPDKLQYTQEIQVKQGRLM GITRRFQVTSGLRQVDQFLGLPYAEAPTGNRRFMPPGAPLPWQGLKIARHLPPVCP QKLPDLSPHGSENMSRARHKHLSRLLPYLRTESEDCLYLNLYVPHEEPQSTPKQYA VLVYLHGESFEWNSGNPYDGSVLSSYGEVIVVTVNYRLGVLGFLRPSIDAHNIANY ALLDQIAALHWIKENIEAFGGDNSRVTLMGHSTGAACVNYLMVSPVASGLFHRAI LMSGSAMSDWAASNQSLQLTMQIAHALECPLNEHVEAEDDDVLLDCLRHRRYQD ILHIPTALTQFSTSLGPIVDGHVIPNQPYKVMGHYTEHFSRYDLLFGITESESYHTLA ALALEEGLRENERDNLLRFYMQSRFDIRPDLALAATLKKYQDMYNNPIKATNLEH RDVVLDILSDARVVGPLLQTGMFHADVNRRNYMYVFGHNSATGPFAHLPHSIMG EELAFIFGAPLAAAGPFPSGNYTVQEKLLSEAVMAYWTNFVKTGNPKAPWKGIFIN SHALEWDRYDLDWPEFNRRAQAYLNIGIPPTVGYKYRQIYMNFWNKELPDELNQI AAIQEQLQKPGQEVITGHMSKYGPRDHGAEDPVRTLKLLLQEPSLAGPTQSGESTE TAAENMYNAPPTFGHVHKMQGGSDFEDLVTSTNSLESGEDHPQAPPETVAKSEAT MQLLIALITIIIVLNLLIYGTFLLRQRRRRAKALPFPAPKLGGTILSYDGANDEELKRC SKSRDGDDSFVLEMTRKSNTYEAIKTGQRSLSCSTVDTHTKVCEWMSSQEAPKSG SFNTATPPPQPILSDGRLIICQDIEVADAALLIPQHMHEPQHYEMLLQRQHSALTEPS DEILQQQYPLRNHSHSHSDPVDMILAADEQVTSFVHADDVDINVTSRDDSDGLEVI PLTSAQQLELLRQRNYPKVLPTEQDLINSSYKRNSLPPQNFNAPLPPPRTISNTLGRR RRDSSNITTSPLQVARDCGGEDEDLKEPQITQNTLIVGPIVPKSPASSLKRVKRMPES SAMTALSGSFQSFEAVPPAHETTPPQGGERTECIYAIRPSPGSCSWAAPNGDLYAQP MKSSSRNSLIPRPTKPAESQSQATPAGPAGESLGSTGMATASRIPQLQRQASGKDLQ ARTATDNTANPLGCQPRTDSTISSGSSASYASTTDSSSSSSTGTVRTELQQFQPAPGR SITTNI
D. ananassae	MNFRSAAFWLFLLTVGGNHRQLSPGGCWGVAAEITQKAASELQQGNELENALKL RATPKQTRFIGDITTTSTDATAAKSIGHQPMRRAKATSTDQNLEHHFKQSVVQLFD SDERYESSEAAITEGGDDDGATATISAANSAEATTTHHPGRRHVVPDKLQYTKEIQ VKQGRLMGITRRFQVTSGLRQVDQFLGLPYAEAPTGNRRFMPPGAPLPWQGLKIA RHLPPVCPQKLPDLSPHGSATMSRGRFKHLTRLLPYLRIESEDCLYLNLYVPHEETQ STPKKYAVLVYLHGESFEWNSGNPYDGSVLASYGEVIVVTVNYRLGVLGFLRPSID AHNIANYALLDQIAALHWIKENIDAFGGDNSRVTLMGHSTGAACVNYLMVSPVAS GLFHRAILMSGSAMSDWAASNQSLQLTMQIAHALDCPLHEHVEAEDDDVLLDCL RHRRYQDILHIPTALTQFSTSLGPIVDGHVIPNQPYKVMGHYTEHFSRYDLLFGITES ESYHTLAALALAEGLRENERDNLLRFYMQSRFDVRPDLALAATLKKYQDMYNNPI KATNLEHRDVVLDILSDARVVGPLLQTGTFHADVNRRNYMYVFGHNSATGPYAH LPHSIMGEELAFIFGAPLAPAGPFPSNNYTVQEKLLSEAVMAYWTNFAKTGNPKAP WKGTFLNSHALEWDRYDLDWPEFNRRSQAYLNIGIPPTVGFKYRQIYMNFWNKEL PDEINQIAAIQEQLHGPSHEVITGHMSKYGPRDHDSEDPVRVLKQLLQEPTLAGPES TETAAENMYNAPPTFGHIHKIQGGSDFELITPPSSFDNGDDQQRSESPETVAKSEITL QLLIALIAIIIALNVIIYGTFWLRQRRRRSKALPFPGPKLGGHILSYDGANDEELKRCS KSRDGDDSFVLEVTRKSNTYEAIKTGQRSLSCSTVDTHTKVCEWMSGQEAPKAGS FATATPPPHQPILADGRLIICQDIEVADAALLIPQHMHEPQHYEMLLQRQHSALTEP SEDVLQPQYPTRSHAHSHSDPVDMILAADEQVTSFVHADDVDINVTSRDDPGVLEI IPLTAAQQLELLRQRNYPKVLPTEQDFLDTSYKRNSLPPQNYNAPQPPPRTISSTLG RRRRDSSNVTTSPLQLAQDISADDEDNLREPQITQNTLIVGPIVPKSPATSLKRRKDK EPAAAAMTALSGSFQSFESVPPAHETTPPQGERTECIYAIQASPASPTTPVSPGSSSW SAPNGDLYAQPVKSPSRTSLIPRPTHSLQVATSQAPQTGDGVPGAIPSRIPQLHRQAS GKDLQATLQAAFTENTANPSGCQQRTDSTISSGSSASNDSSTSSSTGSSSSSSTGTVR TELQQFQPLPGRSITTNI
D. grimshawi	MKFRLGSFWLFLLTVGGNEKQLSADCCHGLATWGIPTAAASTAPLKQKSIEQSDSR

	NRYTHNGNGLENATTLHSTAKQTPTESPQVIGDIITTTPQTTATTARTETQPDATTV TAKSIVELLFRQASRPKSLELSLKEHMATLIGSEQHERGAATTAATPDNDTSSSGTS THHPERRHVVPDKLQYTKELQIKQGRLMGITRRFQVTSGLRDVDQYLGLPYAEAPI GSRRFMPPGAPLPWQGLKIARHLPPVCPQKLPDMSGQNSKSISRARYKHLLRLMPY LKTESEDCLYLNLYVPHEEPLTTSKPHAVLVYLHGESFEWNSGNAYDGSVLASYG EVIVVTVNYRLGVLGFMRPGINAHNIANYALLDQIAALHWIKENIGSFGGDNTRVT LMGHSTGAACVNYLMVSPVASGLFHRAILMSGSAMSDWAASNQSLQLTMQIAQA LGCPLSDHKDEDALLDCLRQHRYQDILHIPTSLQPFSTSLGPIVDGHVIPNQPYKVM GHYTEHFSRYDLLFGITESESYHTLAALALEEGLRENERDNLLHFYMQSRFDVRPD LALAATLKKYQDMYNNPIKATNLEHCDVVLDILSDASVVGPLLQTGMFHADVNR RNYMYVFGHNSAMGPYANLPHSIMGEELPFVFGAPLAPVGPYPSHNYSVQEKLLS EAVMAYWTNFVKTGNPKAPWKGTFLNSHLLEWDRYDLDWPEFNKRAQAYLNIGI PPTVGYKYRQIYMNFWNKELPDELNQIAAIQQKGQQQQQQQQQQFPGNGVITGH MSKYGARDKGAEDPVRTLKLLLQKPLPSKGSEQAETAAENMYNAPPTFGHVHRQ QQQQQPAVEGVAISGEDPTQTSQHVSDYGEYGATVAPTEPIAKSETTLQFLIALIAIF IVLNVVIYATFLFRQRRKQNTSLQRKLGGQMLSYDGANDDELKRCSKSRDGDESY VLDIVRKSNTYEAIKTGQRSPINGFAMQRQLSSSTVDTHTKVCEWMSSQQQSPTQK LLKSGSFNTTPAPLPAAQPDNRIIICEDIEVADAALLTPQHMHEMQEGLHYELLLQR QHSALTEPSEATLQPQHSHTHSDPVDMILAADEQVTSFVHADDVDINVTSRDECEL QLQPLTVAQQLALLQQRNYPKVLPTEQDVRQSYKRNSLPPQNFQAPLPPPRTISSTL GRRRRDSSNITTSPLMLAHDNGEDELHEPQITQNTLIVGPIVPKSPAASLKRIKRQTD TAAISKLSGSFQSFESVPPAHEATPPQGQRTECVYAIEASASVSSLPSWSAPNGDVY AQPTKSASRTSLIPRLLQTAVETQQVSTEATAKATTAAAAAMPSRIPQLQRQASSK ELQSPRESLALLDSTDGQSSNASADSTSRTKSSISNDSSWRSSSASSSSSSTGTVRTE LQQYQQPSKTIATQI
D. pseudoobscura	MKFRLAAFWLFLLTVGSGAAAESIAQRDGNSDSDKGNELENALKLRATSKRKQTQ VIGDITTRAHQSPEDATTTLATASKSIGHQAVRRAKDARNLEQHLKQSVARLFTSD EESNESDSGPGSTLVTDATATATATATATTVTTHHPGRRHVVADKLQYTKEIQIKQ GRLMGITRRFQVTTGLREVDQYLGLPYAEAPTGSRRFMPPGAPLPWQGLKIARHLP PVCPQKLPDLTSHGSVNMSRARHKYLSRLLPYLRSEGEDCLYLNLYVPHEESLVAP KQYAVLVYLHGESFEWNSGNAYDGSVLSSYGEVIVVTVNYRLGVLGFLRPSIDAH NIANYALLDQIAALHWIKENIGSFNGDNTRVTLMGHSTGAACVNYLMVSPVASGL FHRAILMSGSAMSDWAASNQSLQLTMQIAHALDCPLSDHFDADDEDALLDCLRQR RYQDILHIPTSLQQFSTSLGPIVDGHVIPNQPYKVMGHYTEHFSRYDLLFGITESESY HTLAALALEEGLRENERDNLLRFYMQSRFDVRPDLALAATLKKYQDMYNNPIKAT NLEHRDVVLDILSDARVVGPLLQTGMFHADVNRRNYMYVFGHNSATGPYANLPH SIMGEELAFIFGAPLAPAGPFPSHNYTMQEKLLSEAVMAYWTNFVKTGNPKAPWK GIFLNSHALEWDRYDLDWPEFNKRAQAYLNIGIPPTVGYKYRQIYMNFWNKELPD ELNQIAALQEQLHPGSEVITGHMSQYGPRDKGAEDPVRTLKLLLQEPSLANSPDST ETAAENMYNAPPTFGHVMHKQQPQQPQREEDLEMTSPTNPMAEDQPETVAKSET TLQLLIALIALIIALNVIIYGTFLMRQRRKRAKMLPFPGPKLNGNILSYDGANDDELK RCSKSRDGDDSYVMDMARKSNTYEAIKTGQRSPINGYAIQRQLSCSTVDTHTKVC EWMSGQEVPKSGSFATPTATTPTSVLPDGRMIICQDIEVADAALLIPQHMHEPQHY EMLLQRQHSALTEPSEDILQQQLQHMPLHRHTHSDPVDMILAADEQVTSFVHADD VDINVTSRDENCVLEVVPLTAAQQLELLRQRNYPKVLPTAQDLRDSRGGNYKRNS LPPQNFSAPLPPPRTISSTLGRRRRDSSNITTSPLQLAQDLAGEEEPGEPQITENTLIFG PIVPKSPATSLKRQKRTPEAGAPAAGAMTALSGSFQSFESVPPAHETTPPQGGERTE CIYSIEASPGTCPGNDGTSWSAPNGDLYAQPMKSPSRTSLIPRLTQQQPTPTPTPTPT PTPTPTPSPHHHHQGAAEGEETSAVPGGASSMATTVSPTRIPQLLQRQASGKDLRIA NTASPLASQGQQQQQQQQRTDSTISSGSSASAASSTESSTSSSSSSTNTVRTELQQH QHQHQPGKSITTNI
D. virilis	MKFHLGAFWLFLLTVGGSQKQLSDNCCHGLATWGIPAVAAATAALPKQKSIQSPP NFRNGSKHNGNELENATKFRVAAMQTPAEPSPVIGDITTTTTTTTTTTRTTWPPAQ

	PDVTTLTAKSIVQLTFKQAKAHNLELSFKERMSQLFASEESESSLTTTITTTTTTSSDI DNSSGSGTSTHHPERRHVVPDKLQYTKELQIKQGRLMGITRRFLVTSGLREVDQYL GLPYAEAPIGSRRFMPPGAPLPWQGLKIARHLPPVCPQKLPDVSGPSSVNMSRGRY KHLMRLMPYLKTESEDCLYLNLYVPHEEPLAKPKLHAVLVYLHGESFEWNSGNA YDGSVLASYGEVIVVTVNYRLGVLGFMRPGIDAHNIANYALLDQIAALHWIKENIG SFGGDNTRVTLMGHSTGAACVNYLMVSPVASGLFHRAILMSGSAMSDWAASNQS LQLTMQIANALDCPLGDHEEDDALLDCLRQRRYQDILHIPTSFQQFSTSLGPIVDGH VIPNQPYKVMGHYTEHFSRYDLLFGITESESYHTLAALALEEGLRENERDNLLRFY MQSRFDVRPDLALAATLKKYQDMYNNPIKATNLEHRDVVLDILSDARVVGPLLQT GMFHADVNRRNYMYVFGHNSATGPYAHLPHSIMGEELAFVFGAPLAPAGPFSSHN YSVQEKLLSEAVMAYWTNFVKTGNPKAPWKGTFLNSHLLEWDRYDLDWPEFNK RAQAYLNIGIPPTVGYKYRQIYMNFWNKELPDELNQIAAIQVRGQQQQLNGNEVIT GHMSKYGARDNGAEDPVRTLKLLLQEPLATGSEQLETAAENMYNAPPTFGHVHR QQQQQQQQQQQGIGTNGAEGVAISGEDTTHSNDYGEYGASSSTATPAEPIAKSETT LQLLIALIAIFIVLNVAIYATFLLRQRRKRNTSLQRKLGGNILTYDGANDDELKRCS KSRDGDDSFTLDMVRKSNTYEAIKTGQRSPINGYAMHRQLSSSTVDTHTKVCEWM SAQHQHTNSGHKLPKSGSFNTTPPPPPAVQSDSRIIICQDIEVADAALLTPQHMHET LEGMHYELLLQRQHSALTEPSEATLQPQHGHTYAHAHSHSDPVDMILAADEQITSF VHADDVDINVTSRDEDALVLQPLSAAQQLELLRQRNYPKVLPTAQDLRTSHSYKR NSLPPQNFTAPLPPPRTISSTLGRRRRDSSNITTSPLMLAQDSGEEEPREPQITQNTLI VGPIVPKSPAASLKRIKHTGVATALSTLGGSFQSFESVPPAHEATPPQGQRTECVYA IAASASVSSVPGISSPSSWSAPNGDLYAQPTKSASRTSLIPRPLQTAPAAAAAEAPAA TQHVSTAADSATPSRIPQLQRQASGKELSPVDGQQTVDATTPTSNSRVGSTISSDSS ASAASWRNSTTSSSSSSTGTVRTELQHCQQSGRTIATHI
D. willistoni	MKFRLAAFWLFLLTIGDSRHHFSHGPGVAAAAMQQKSISQQQRNKGNELENAVQ LRETSKQMQVIGDITTTQPDASSMLMGKSIAQQTFRHAKTNTPPNLELNFKQTVAK LLSSDEHDSEFDPEANDDEADEGTAATAGTESFTTHHPERRHVIPDKLQYTKEIQIK QGRLMGITRRFQVTSGLREVDQYLGLPYAEAPTANRRFMPPGAPLPWQGLKIARH LPPVCPQNLPDISPQASGSMSRGRYRYLSRLMPYLRTESEDCLYLNLYVPHEAAEQ LAAPKLHPVLVYLHGESFEWNSGNAYDGSVLASYGEVIVVTVNYRLGVLGFLRPSI DAHNIANYALLDQIAALHWIKENIGSFGGDSSRVTLMGHSTGAACVNYLMVSPVA SGLFHRAILMSGSSMSDWAASNQSLQLTMQIAHNLDCPLGDHIDDDALLECLRQR RYQDILHIPTTLPQFSTSLGPIVDGHVIPNQPYKVMGQYAEHFSRYDLLFGITESESY HTLAALALEEGLRENERDNLLRFYMQSRFDVRPDLALAATLKKYQDMYNNPIKAT NLEHRDVVLDILSDARVVGPLLQTGMFHADVNRRNYMYVFGHNSAMGAYANLP HSIMGEELAFIFGAPLAPAGPFPSHNYSVQEKLLSEAVMAYWTNFVKTGNPKAPW KGTFLNSHALEWDRYDLDWPEFNKRAQAYLNIGIPPSVGYKYRQIYMNFWNKELP DELNQIAAIQQQQHHQQQQQQESNVVAGSNVITGHMSNYGSRDNGAGDPVRTLK MLMGVVGEPTEAAAENMYNAPPTFGHVHKQQGEGGEEMVTGIPFSPGSNDVKTE SSSSTETVAKSEITMQLLIALIAMIIILNVIIYGTFLIRRRRAKTSPLKRKLGGNILSYD GANDDELKRCSKSRDGDDSYILDVVRKSNTYEAIKTSQRQLSCSTVDTHTKVCEW MSGHPSSHNHTLPKSGSFATNTPSPPPAMANADGSRIIICQDMEVSDLALLTPQHM HDMQHYELLFQRQNSALTEPSEDVLQPATHHQHHHHHQHSHSDPVDMILAADEQ VTSFVHADDVDINVTSRDDRLDNDAVPLTAAQQLELLRQRNYPKVLPTAHDLAAK GVLAVNFKRNSLPPQNFNPPVPPPRSTSQTLGRHRRDSTNVTTSPLQLARDCGEDEP GEPLITQNTLIFGPIVPKSPATSLKRPKPRTLPGSETELPTAMNALSGSFQSFESVPPP VHEDTPPQGERTECVYSIEPTGSGPSSPGAWSAPNGDLYASPIKSPSRTSLIPRPMQT AAAPAKTTTEPQSPTCVGDAEAVLNPNPTRIPQLQRQASGKDLYQPTLRTNSTISSG SSASAASGTSATSATSSSSSSTGTVRTELQKSPSQQQQSSKCITTKI
Lutzomyia longipalpis	MLQLRFPMMILFLALCVSLSCGMIDASTENLASSRQNSDTTKTSRLAGRELLVNIPT QEKVRISPTQKTRRQSQRNMTLTRQIAIKQGILRGVVRFMHPQSGLRDVDQFLGVP YAEPPINSKRFMPPGAPLPWSDVRIFSELPPVCPQNLPNLHLANGTLSKGRYNQLKR LLPYLKTEREDCLYLNIYMPTWEGSIFQAKFPVIVYIHGESFEWNSGNPYDGSILAS YGRVIVVTVNYRLGILGFMRPGVGEEIVSNFGLLDQIAALQWIKENIEAFGGDSSSV

	TLMGHGTGAACINFLMVSPVAKGLFHRAILMSGSALSDWATSKHPTQFTMQVAQS LNCPITDDELVACLRRKRYTEMLRAKVSSPKFSTIFGPLIDGLVIPNEPATIMSQSTD TFSKYDLLFGMTEVESYHMLNAIELMYGLLENERDTLLRFYLQNRFEIRPDLALAA TLKEYSNIYATEKTQNANFHRDKILEIFSDARVAGPLVQTGLFHAKSNPRSFMYVF AHNSEAGEYAYLSQSIIGEELAYVFGAPLSPSGPFQPFYNSQEKLLSEAVMKYWTN FAKTGNPKAPWRDKFLNLNPIEWARFDIDWPEYNQINQSYLHLNSHPVVGHQYRH TFMRFWNQNLPSELHKLTTARTNIHMMPPSIPSAEGPRREIYGNFSTHPRHFGGGEV RSRPDDMMYAKTENPIAILRLMNKNTRPSHGVTVSEEKMSTEAANVAPFDPDSYV PNSANLTLNLLIIIVVILSILIVIFGYILKRSYKRSSPTEKILKNLKRSTPPKMFYRSDN DLSEMSNNSYDMVRIERGSKVDSIHSDGINYPTTVIKAPWRTLPKRCSFEDPDNQS DCFIKPGQVSVAIDATPQARSGSVLKQEPIEVTKSRILIAPETKIICRELDNPPLTVYY GDDLCTEEEYPRKPEENHYNTMGNGHVRNFSDSQKGGEVSAVIVRGTNSLDSMK MRRRSLPAQNQLDSALGWTKVPPMPPPRTVSTLSRRHNPPVPEENLPHPPQHMLA TLPPSSDLEDEVITSNTLIVGPLIPTKGTKGKSSDASDDGDQDEATLSGVASTTKYN ALSRSKKKSTASKNEPNIIIKPQITRSTSEKANKFNARRMPAAANSAGTTETPD
Phlebotomus papatasi	MIDASTENLASTVQNSDTTKTSRGTNREINLQATERVAKSSPTQKSRRQSQRNITLT RQIAIKQGILRGVVRHMHPHSGLRDVDQFLGVPYAEPPINNKRFMPPGAPLPWSDV RVFSELPPVCPQNLPNLQATNGTLSKGRYNQLKRLLPYLKSEREDCLYLNIYMPTW EGSIFQAKFPVIVFIHGESFEWNSGNPYDGSVLASYGRVIVITVNYRLGILGFMRPG VSEEVVSNFGLLDQIAALQWIKENIEAFGGDSSSVTLMGHSTGAACINFLMVSPVA KGLFHRAILMSGSALSDWATSKHPTQFTMQVAQSVHCPIADDELVACLRKKRYSE MLRAKVSSPKFSTIFGPLIDGLVIPNEPASIMSQSTDTFSKYDLLFGMTEVESYHML NAIELMYGLLENERDTLLRFYLQNRFEIHKVLEIFSDARVAGPLVQTGLFHAKSNPR SFMYVFAHNSEAGEYAYSIIGEELAYVFGAPLSSSGPFQPFYNSQERLLSEAVMKY WTNFAKTGNPKAPWRDKFLNGNPIEWARFDIDWPEYNQINQSYLHLNSHPAIGHQ YRHIFIKFWNQNLPSELHKLTTSRSSVHMIPPNIPSPDVPRREIYGNFSTHTKHFGGG GEIKTHPDDRMYSKTDNPMTILRLLNKSPRPTPEIHSFEDKVSTEVGDVSPFDPESYI PNSANLTLNLLIIIVVILSILIIIFGYILKRSYKRSSPTEKMLKNLKRSTPPKMFYRSDN DLSEMSNNSYDMVRIERGSKIDSIHSDGINYPTTVIKAPWRTLPKRCSFEDPDNQSD CFMKPGQVSVAIDATPQARSGSVLKQEPIEVTKSRVMLAPETKIICQELDNPRISAY YGNDLCDEDDYVRKPPDSGFSTLGNGHVRNFSDTQKCGELPSTFARKTNSLDSMK MRRRSLPAQNQLDFAMGWTKVPPVPPPRTVSTLTRRQNPTVPEEILPHPPQHMLAT LPPSSNLEDEVITSNTLIVGPLIPSKGPKGKSSDASDEGESAEATLSGTSSTTKYNALS RSKKKSTSRNEPNIIIKPQISRSTSEKANKFSARRIPASSTTTNPAEANETNI
Aedes aegypti	MKLFVVLINLVLCFSLAYQQVETRTDSDEAFDVYEDLVSDRSENPKKSIAALLAHD GSEIFPPMSNDGPFGDYRIQSHQQYGSSAYRKQHAQRSQSSSFDPHDGTPYTREVA VKQGRIKGIVRVMHPQSGLKSVDQYLGIPYAEAPVGSRRFMPPSAPIPWTALKMAI KMSPVCPQNLPTLNNVNNNYSKGRYDQLKRLLPYLKVESEDCLYLNLYVPSYDGI GPQAKYPVIVYIHGESYEWNSGNPYDGSILASYGRVIVVTLNFRLGILGFMKPGISD HTTSNFGLLDQIAALQWIKENIGAFGGDNKLVTVMGHGTGAACVNFLMVSPVAK GLFHRAILLSGSALSDWALTQHPLQSTMQVLQGLNCPLNGDNDEVATCLRRKRYS EILNVKIASPQFSTRFGPIVDGLVIPNTPHKVMGQYSDIFSGYDLLYGMTELESYNIL NAVALTYGLLENERDNLLRFYMQNRFEIRPDLALAATLNEYTGILTDPNKSLADIH RDTLLDILSDARVAAPMLQTGLYLAKVNPKCYMYVFGHNSEAGEYGRLSQSVVG EDLAYIFGAPLGPVGPFQTHYNARERLFSEAVMKYFSNFAQTGNPKAPWKDLFLN MNPEDWRYYDVDWPEFNSVNQSFLHMGLTPVVSRHYRQQYMKFWNQKLPEELK RITTSKPYSPYSELISPPDRRMTTSSPEFVTGHINLFPIKVDIEKPTEDSFKALVHRMK DPGMYNAPASSIGIPAEAPVKEIPKPENVEKDMENSEMMKTEASLTLLITVAIVFLIF NIFVIVGYIIKRNVGNRKMKRKFDDNILESAASMNEKRSKLNDTEDSYILDLMRKS NPYEPVAKANNFTLPRQFSGETMDPHTKVCDWISQDYGTVCTKQIECDGNGSCNK DNLPLPQKVSVAIDATPQARSNSVLRQEPIEITKAKSFECASPEDYSESVSQDSPCDS DIRQQCSSSGSSYSYNDCDYKVEPKHFKSDSNPIYNNYPYQNEEITSFIEPGDINVTS RDKSEEKDPLSPEEALKVIQRRNFPKVLPDYPNGVAGITASMKRRSLPPQAFLLANS HSLRRDSSRLVPAPPPRISSTLGRRSSKNRTSNNFLSSPPVMAEEPPIEEEPPITLNTLH

	VGPLLPKHHENTYMTMSRQNSAETEQTTSTEEPAVDIICIEHKPESHYSYIKPPSTIK PPTSFKSSPPKELNRTESRAELQQASTYGSTPRLPITRSGSSIMHREMSSDSSATDTTS GSTGTIKML
Anopheles gambiae	MTILLLLLLLTLSVQPGQSQVETILDSSEYDDLGLPTVTERKGQMAGFLASDGSDL YPSISMSNDGPSVPYQQRPSNYATNPFIARAEQYRKAHASLRPQPQGSSAQQTQQQ QQQQQQQGYDPRDGTPYTRDIAVKQGILRGSVRVMHPQSGLKNVDQFLGIPYAEA PVGSRRFMPPSAPIPWNGLKMATKLSPVCPQNLPSLNNANNNYSKGRYDQIKRLLP YLKVESEDCLYLNLYVPNGIGPQTKYPVIVYIHGESYEWNSGNPYDGSILASYGQVI VVTLNFRLGILGFMKPGISDHTTSNFGLLDQIAALQWIKENIGAFGGDAKLVTVMG QGTGAACVNFLMVSPVAKGLFHRAILMSGSALSDWALTQHPLQSTMQVLQGLNC PLNGENDEVTACLRRKRYSEILGVKTASPQFSTRFGPIVDGLVIPNMPHKVMGQYD LLYGMTELESYHILNAVALTYGLLENERDNLLRFYMQNRFEIRPDLALAATLREYT DIYMDPNKALADEHRDNLLEILSDARVAAPMVQTGLYLSKVNPKCYMYVFGHNS EAGEYGRLSQSVVGEDLAYVFGAPLGQVGPFQHHYNARERLFSEAVMKYFSNFA KTGNPKAPWKDLFLNLNPEDWSYYDVDWPEYNSINQSYLHMGITPVVGHRYRQK YMKFWNEELPEELKKITSSKAYAPYSDFFSPPGAGGKQGGGVSAVVGRMRGGTTP HPDYPTGHINLYPIHVDIERPTEDPFRELLYQMKDPLAGPGMYNAPATSSAAAIPEK GASPAQLAAGVLEHPKHPKQLQQSGKDGPFGVTDEQELDGGATEIMKSESTLTILI AVAIVFLLFNIVVIVGYLVRRHLGRPGVGGVGVGGVAGGRKVKRKYDDTMLESA AAVVVSGVGLGEKEDGCNQLHGHHHHHHLPHGAYLLDEAATMMLRKSNSYEPV AKQPFGGELHGPVSIGVDAVDAHTKVCDWMVASAGCRMAPAVPPHKISVAIDAT PQARGNSVLRQEPIEITKAKSFEYGATDEVDCAVSGCPGEGRMGEGNASGTGSSSS GTSSSCSSSATTDELKQTMPGAGAGLYYNCPDYPVRRFGATGEGGEEEVTSFIEQD TPAGDINVTSRDKSTEKDPLSPEEALRVIQRRNFPKVLPDHPRTGGTGGPTTVGTSM KRRSLPPQSLYGALPHGNCHSLRREAAGAGRLIPAPPPRICSTLGRPAAVRQARASN NFISSPPIVAEEPPVEEEPPIALNTLHVGPLLPTHQESTYMTMSRQNSLGSECEPGSSE ADDDRPSEEQPPVDIICIEHKPENHYSYVRPVGLMRTPSSFKCPEAVSRTEGGEARS AGDGETGQPTPIDAATQPLQPVRDKFSSDSSATDTTSGSTGTIKK
Culex quinquefasciatus	MKVFVVLINLVLCFSLAYQQVETRTDSDESWGPGQYDEGGLYGGHRNEHSRKSV AALLAHDGSEIFPPMSNDGPQYGASSKILQQNQQQYGVPSYRKQPAPRSQSPSFDP RDGTPYTREVAVKQGRLKGIVRVMHPQSGLKNVDQYLGIPYAEAPVGSRRFMPPG APVPWTGLKMAIKMSPVCPQNLPTLNNVESEDCLYLNLYVPSYDGIGPHAKYPVIV YIHGESYEWNSGNPYDGSILASYGRVIVVTLNFRLGILGFMKPGISDHTTSNFGLLD QIAALQWIKENIGAFGGDNKLVTVMGHGTGAACVNFLMVSPVAKGLFHRAILLSG SALSDWALTQHPLQSTMQVLQGLNCPLNGDNDEVAACLRRKRYSEILNVKIASPQ FSTRFGPIVDGLVIPNTPHKVMGQYSDIFSGYDLLYGMTELESYNVLNAVALTYGL LENERDNLLRFYMQNRFEIRPDLALAATLNEYTGILTDPNKSLADIHRDTLLDILSD ARVAAPMLQTGLYLAKVNPKCYMYVFAHNSEAGEYGRLSQSVVGEDLAYVFGA PLGPVGPFQTHYNARERLFSEAVMKYFSNFAQTGNPKAPWKDLFLNLNPEDWRYF DVDWPEFNSVNQSFLHLGLTPVVSRHYRQQYMKFWNQKLPEELKRITTSKPYSPY SEFISPPDRRMTTSNPEFVTGHINLFPIKVDIERPTEDPFKALVYKMKDAQVGAPAP AAVGMPAEAPAKEAPKPANVEQESENSEMMKTEASLTLLIAVAIVFLLFNIFVIVG YVIKRNVGNKKVKRKFDDNILESAASMNEKRSKLNETEESYILDMMRKTNPYEAV VKGNNFTLQRQFSEDTVDSHTKVCDWIAQDYGAGCPNQTQCEGLGSIPSCNAATL NVPHKISVAIDATPQARSNSVLRQEPIEITKAKSFEYASPNDYDESVSQDSPTDSDM RQQCSSSGSSYSYNDCDYKIEPLLSKSDANPVYNNYPYQTTEDVTSFIEPGDINVTS REKSQEKDPLSPEEALKFAEGRQSPHSGTTTPNLIDPGKEIFQAPIQQQFP
Daphnia pulex	MLRFTTLVWLVAIFACGFIVCVAGQMRPSLNARIVRTKQGSVKGVLVIPSNRELQP VEAFLGLPYASPPVGPLRFMSPVSPLPWNGVRLMDKYAPACPQTLPDVSNEREAL RFVTRGRLQYLRRLLPYLRNQSEDCLYLNIYAPVTGTTVGGKDVNIKFPVIVFIHGE SYEWNSGNPYDGSILASYGDVVVVTINFRLGVLGFLRPDLRENRVANFGLLDQIAA LQWIQENIAQFGGDRDSVTLLGHGTGAACVNLLLISPVAQSSSGLFHRAILMSGTA LADWAVAENPLRYTLQAAQQVDCPLAERDDELAACLRFKRVTELMSAPLYASPL

	APLVDGIVVPNEPRQSMKIYNELFGRYDVMYGVTQSESFHLLNAATLQFGLTETER NRIIRTYVRNSYASNIDQVTAAILNEYTDWKNPVRDTEEYRDSILEILSDARVVAPV IQMADLHSSIRQRSYFYVFAHQTTNGDYPQNWGSIHGEELAYVFGMPLVGGTNHL SSNYTRAEMLLSEIVITYWANFARTGNPNFPPRQKFLTVANSRDRFEPSYVQWPAY DRHSQKFLNIDLRPKVKDHYRADKMALWSKLIPELMSNGNGKVFTVGNEDDDDD DDGEVLVTGGGEASSVAEPDSSSHSPPAVMGAQPEVYQQPSLGGFEVFPIGPMSKG ASGNSSRPDDGQDLVPQTGGIALSIVIVIGICFLVLNVCACAGVFYQRDRVRFKEML IQRQYKLKSANHDGEGPTSGAQVAAAGARETLLALHRVEEDFPELDGSANGMQS GLPHQASTSTMDPHTKVSQWMAQEITIERCPTPPINNRLAKPAVQDKLGGSDRYDS RGYDMEDPASALFPLLTKSPNHQDIYGLLPVKEENSADAKQPITSGQSPDFINQVH NNGTFLKFGDVGHVGESSPPFADSASSAQTMSRSSTGRRKARSKSQLGSQRSITTK RDVAVGGDGDDDEGNYRVAPDNGGEDDRRASALYGYGAMDTTIRRLNLPKVLP DLPHQDVPAEATAPSGVDLTGPKTSSSNLVPSDQHSPIYTIPYAGLPTGVDSGERST PISRNERHPRQTMESPPASTVPLSADTKKQLSKSRRKYAGSAGGPANTSGGSSEAV LLSSPTSTIVKTNLPQTVVVAPHPRANRVAASSPTAGLLHQQQQPEPCLVIRPGPRQ TCQSTSTASTAGGPAGVSGYDSATNDNEFYPSATADVILRRPRQQDASTNNPTRPA NRNSRSWYAQYSQSFISQSIDQESDKNDN

Figure 5. Neuroligin 1 Homologue Alignment

```
CLUSTAL 2.1 Multiple Sequence Alignments


Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: Dmel        1354 aa
Sequence 2: Dana        1370 aa
Sequence 3: Dgri        1414 aa
Sequence 4: Dpso        1413 aa
Sequence 5: Dvir        1438 aa
Sequence 6: Dwil        1386 aa
Sequence 7: Llon        1127 aa
Sequence 8: Ppap        1075 aa
Sequence 9: Aaeg        1252 aa
Sequence 10: Agam        1380 aa
Sequence 11: Cqui        1052 aa
Sequence 12: Dpul        1255 aa
```

---

clustalw.aln

```
CLUSTAL 2.1 multiple sequence alignment


Llon            MLQLRFPMMILFLALCVSLSCGMIDASTENLAS--------------------------
Ppap            ----------------------MIDASTENLAS--------------------------
Aaeg            ---MKLFVVLINLVLCFSLAYQQVETRTDSDE----------------------------
Cqui            ---MKVFVVLINLVLCFSLAYQQVETRTDSDESWGPG-----------------------
Agam            ---MTILLLLLLLTLSVQPGQSQVETILDSSE----------------------------
Dmel            ---MKFRLAAFWLFLLTVGGNHKLLSHVSPMGVAAEK-----------------------
Dana            ---MNFRSAAFWLFLLTVGGNHRQLSPGGCWGVAAEITQ--------------------K
Dpso            ---MKFRLAAFWLFLLTVGSGAAAESIAQRDGNSDSD-----------------------
Dgri            ---MKFRLGSFWLFLLTVGGNEKQLSADCCHGLATWGIPTAAASTAPLKQKSIEQS--DS
Dvir            ---MKFHLGAFWLFLLTVGGSQKQLSDNCCHGLATWGIPAVAAATAALPKQKSIQSPPNF
Dwil            ---MKFRLAAFWLFLLTIGDSRHHFSHGPGVAAAAMQQKSISQQ----------------
Dpul            ---MLRFTTLVWLVAIFACGFIVCVAG---------------------------------
                                        :

Llon            -------------------------------------------------SRQNSDTTKTSRLA
Ppap            -------------------------------------------------TVQNSDTTKTSRGT
Aaeg            ----------------------------------AFDVYEDLVSDRSENPKKSIAALLA
Cqui            ----------------------------------QYDEGGLYGGHRNEHSRKSVAALLA
Agam            ---------------------------------YDDLGLPTVTERKGQMAGFLA
Dmel            --QPLHKGNELENAMKLREAPK--QSRVIG----DITTTIQ--------PDSDPGRSIGH
Dana            AASELQQGNELENALKLRATPK--QTRFIG----DITTTS---------TDATAAKSIGH
Dpso            ------KGNELENALKLRATSKRKQTQVIG----DITTRAHQSPEDATTTLATASKSIGH
Dgri            RNRYTHNGNGLENATTLHSTAKQTPTESPQVIGDIITTTPQTTATTARTET--QPDATTV
Dvir            RNGSKHNGNELENATKFRVAAMQTPAEPSPVIGDITTTTTTTTTTTRTTWPPAQPDVTTL
Dwil            ---QRNKGNELENAVQLRETSKQMQVIG-----------------DITTTQPDASSMLMG
Dpul            ---------------------------------------------------------------


Llon            G-------------------------------------------------RELLVNIPTQEK
Ppap            N-------------------------------------------------REINLQATERVA
Aaeg            HDGSEIFP--PMSNDGPFG-----------------------DYRIQSHQQYGSSAYRK
Cqui            HDGSEIFP--PMSNDGPQYGAS--------------------SKILQQNQQQYGVPSYRK
Agam            SDGSDLYPSISMSNDGPSVPYQQRPSNYATNPFIARAEQYRKAHASLRPQPQGSSAQQTQ
Dmel            QALRRAKAP--PPSLELHFRQNLNKLFAGDETTHVGHTNATDN-TTAAVLATDEGSSEHE
Dana            QPMRRAKATSTDQNLEHHFKQSVVQLFDSDERYESSEAAITEGGDDDGATATISAANSAE
Dpso            QAVRRAKDAR---NLEQHLKQSVARLFTSDEESNES-DSGPGSTLVTDATATATATATAT
Dgri            TAKSIVELLFRQASRPKSLELSLKEHMATLIGSEQHERGAATTAATPDND------TSSS
Dvir            TAKSIVQLTFKQAKAHN-LELSFKERMSQLFASEESESSLTTTITTTTTTSSDIDNSSGS
Dwil            KSIAQQTFRHAKTNTPPNLELNFKQTVAKLLSSDEHDSEFDPEANDDEADEGTAATAGTE
Dpul            ---------------------------------------------------------------


Llon            VRISPTQKTRRQSQRNMTLTRQIAIKQGILRGVVRFMHPQSGLRDVDQFLGVPYAEPPIN
```

```
Ppap            K-SSPTQKSRRQSQRNITLTRQIAIKQGILRGVVRHMHPHSGLRDVDQFLGVPYAEPPIN
Aaeg            QHAQRSQSSSFDPHDGTPYTREVAVKQGRIKGIVRVMHPQSGLKSVDQYLGIPYAEAPVG
Cqui            QPAPRSQSPSFDPRDGTPYTREVAVKQGRLKGIVRVMHPQSGLKNVDQYLGIPYAEAPVG
Agam            QQQQQQQQQGYDPRDGTPYTRDIAVKQGILRGSVRVMHPQSGLKNVDQFLGIPYAEAPVG
Dmel            PSTTHHPERRHVVPDKLQYTQEIQVKQGRLMGITRRFQVTSGLRQVDQFLGLPYAEAPTG
Dana            ATTTHHPGRRHVVPDKLQYTKEIQVKQGRLMGITRRFQVTSGLRQVDQFLGLPYAEAPTG
Dpso            TVTTHHPGRRHVVADKLQYTKEIQIKQGRLMGITRRFQVTTGLREVDQYLGLPYAEAPTG
Dgri            GTSTHHPERRHVVPDKLQYTKELQIKQGRLMGITRRFQVTSGLRDVDQYLGLPYAEAPIG
Dvir            GTSTHHPERRHVVPDKLQYTKELQIKQGRLMGITRRFLVTSGLREVDQYLGLPYAEAPIG
Dwil            SFTTHHPERRHVIPDKLQYTKEIQIKQGRLMGITRRFQVTSGLREVDQYLGLPYAEAPTA
Dpul            ------------QMRPSLNARIVRTKQGSVKGVLVIPSNRE-LQPVEAFLGLPYASPPVG
                                    :: :  *** : *          *: *: :**:***..*

Llon            SKRFMPPGAPLPWSDVRIFSELPPVCPQNLPNLHL---ANGTLSKGRYNQLKRLLPYLKT
Ppap            NKRFMPPGAPLPWSDVRVFSELPPVCPQNLPNLQA---TNGTLSKGRYNQLKRLLPYLKS
Aaeg            SRRFMPPSAPIPWTALKMAIKMSPVCPQNLPTLNN---VNNNYSKGRYDQLKRLLPYLKV
Cqui            SRRFMPPGAPVPWTGLKMAIKMSPVCPQNLPTLN-----------------------NV
Agam            SRRFMPPSAPIPWNGLKMATKLSPVCPQNLPSLNN---ANNNYSKGRYDQIKRLLPYLKV
Dmel            NRRFMPPGAPLPWQGLKIARHLPPVCPQKLPDLSPH--GSENMSRARHKHLSRLLPYLRT
Dana            NRRFMPPGAPLPWQGLKIARHLPPVCPQKLPDLSPH--GSATMSRGRFKHLTRLLPYLRI
Dpso            SRRFMPPGAPLPWQGLKIARHLPPVCPQKLPDLTSH--GSVNMSRARHKYLSRLLPYLRS
Dgri            SRRFMPPGAPLPWQGLKIARHLPPVCPQKLPDMSGQ--NSKSISRARYKHLLRLMPYLKT
Dvir            SRRFMPPGAPLPWQGLKIARHLPPVCPQKLPDVSGP--SSVNMSRGRYKHLMRLMPYLKT
Dwil            NRRFMPPGAPLPWQGLKIARHLPPVCPQNLPDISPQ--ASGSMSRGRYRYLSRLMPYLRT
Dpul            PLRFMSPVSPLPWNGVRLMDKYAPACPQTLPDVSNEREALRFVTRGRLQYLRRLLPYLRN
                  ***.* :*:**  :::  . .*.***.** :                          .

Llon            EREDCLYLNIYMPT----WEGSIFQAKFPVIVYIHGESFEWNSGNPYDGSILASYGRVIV
Ppap            EREDCLYLNIYMPT----WEGSIFQAKFPVIVFIHGESFEWNSGNPYDGSVLASYGRVIV
Aaeg            ESEDCLYLNLYVPS----YDGIGPQAKYPVIVYIHGESYEWNSGNPYDGSILASYGRVIV
Cqui            ESEDCLYLNLYVPS----YDGIGPHAKYPVIVYIHGESYEWNSGNPYDGSILASYGRVIV
Agam            ESEDCLYLNLYVP------NGIGPQTKYPVIVYIHGESYEWNSGNPYDGSILASYGQVIV
Dmel            ESEDCLYLNLYVP----HEEPQSTPKQYAVLVYLHGESFEWNSGNPYDGSVLSSYGEVIV
Dana            ESEDCLYLNLYVP----HEETQSTPKKYAVLVYLHGESFEWNSGNPYDGSVLASYGEVIV
Dpso            EGEDCLYLNLYVP----HEESLVAPKQYAVLVYLHGESFEWNSGNAYDGSVLSSYGEVIV
Dgri            ESEDCLYLNLYVP----HEEPLTTSKPHAVLVYLHGESFEWNSGNAYDGSVLASYGEVIV
Dvir            ESEDCLYLNLYVP----HEEPLAKPKLHAVLVYLHGESFEWNSGNAYDGSVLASYGEVIV
Dwil            ESEDCLYLNLYVPHE--AAEQLAAPKLHPVLVYLHGESFEWNSGNAYDGSVLASYGEVIV
Dpul            QSEDCLYLNIYAPVTGTTVGGKDVNIKFPVIVFIHGESYEWNSGNPYDGSILASYGDVVV
                : *******:* *              ..*:*::****:******.****:*:*** *:*

Llon            VTVNYRLGILGFMRPGVGEEIVSNFGLLDQIAALQWIKENIEAFGGDSSSVTLMGHGTGA
Ppap            ITVNYRLGILGFMRPGVSEEVVSNFGLLDQIAALQWIKENIEAFGGDSSSVTLMGHSTGA
Aaeg            VTLNFRLGILGFMKPGISDHTTSNFGLLDQIAALQWIKENIGAFGGDNKLVTVMGHGTGA
Cqui            VTLNFRLGILGFMKPGISDHTTSNFGLLDQIAALQWIKENIGAFGGDNKLVTVMGHGTGA
Agam            VTLNFRLGILGFMKPGISDHTTSNFGLLDQIAALQWIKENIGAFGGDAKLVTVMGQGTGA
Dmel            VTVNYRLGVLGFLRPSIDAHNIANYALLDQIAALHWIKENIEAFGGDNSRVTLMGHSTGA
Dana            VTVNYRLGVLGFLRPSIDAHNIANYALLDQIAALHWIKENIDAFGGDNSRVTLMGHSTGA
Dpso            VTVNYRLGVLGFLRPSIDAHNIANYALLDQIAALHWIKENIGSFNGDNTRVTLMGHSTGA
Dgri            VTVNYRLGVLGFMRPGINAHNIANYALLDQIAALHWIKENIGSFGGDNTRVTLMGHSTGA
Dvir            VTVNYRLGVLGFMRPGIDAHNIANYALLDQIAALHWIKENIGSFGGDNTRVTLMGHSTGA
Dwil            VTVNYRLGVLGFLRPSIDAHNIANYALLDQIAALHWIKENIGSFGGDSSRVTLMGHSTGA
Dpul            VTINFRLGVLGFLRPDLRENRVANFGLLDQIAALQWIQENIAQFGGDRDSVTLLGHGTGA
                :*:*:***:***::*.:  .  :*:.********:**:***  *.**   **::*:.***

Llon            ACINFLMVSPVAK---GLFHRAILMSGSALSDWATSKHPTQFTMQVAQSLNCPIT-----
Ppap            ACINFLMVSPVAK---GLFHRAILMSGSALSDWATSKHPTQFTMQVAQSVHCPIA-----
Aaeg            ACVNFLMVSPVAK---GLFHRAILLSGSALSDWALTQHPLQSTMQVLQGLNCPLNG----
Cqui            ACVNFLMVSPVAK---GLFHRAILLSGSALSDWALTQHPLQSTMQVLQGLNCPLNG----
Agam            ACVNFLMVSPVAK---GLFHRAILMSGSALSDWALTQHPLQSTMQVLQGLNCPLNG----
Dmel            ACVNYLMVSPVAS---GLFHRAILMSGSAMSDWAASNQSLQLTMQIAHALECPLNEHVEA
Dana            ACVNYLMVSPVAS---GLFHRAILMSGSAMSDWAASNQSLQLTMQIAHALDCPLHEHVEA
Dpso            ACVNYLMVSPVAS---GLFHRAILMSGSAMSDWAASNQSLQLTMQIAHALDCPLSDHFDA
Dgri            ACVNYLMVSPVAS---GLFHRAILMSGSAMSDWAASNQSLQLTMQIAQALGCPLSDHKDE
Dvir            ACVNYLMVSPVAS---GLFHRAILMSGSAMSDWAASNQSLQLTMQIANALDCPLGDHEED
Dwil            ACVNYLMVSPVAS---GLFHRAILMSGSSMSDWAASNQSLQLTMQIAHNLDCPLGDHIDD
Dpul            ACVNLLLISPVAQSSSGLFHRAILMSGTALADWAVAENPLRYTLQAAQQVDCPLAER---
                **:* *::****.   ********:**::::*** :::. : *:*  : : **:

Llon            --DDELVACLRRKRYTEMLRAKVSSPKFSTIFGPLIDGLVIPNEPATIMSQSTDTFSKYD
Ppap            --DDELVACLRKKRYSEMLRAKVSSPKFSTIFGPLIDGLVIPNEPASIMSQSTDTFSKYD
```

```
Aaeg            -DNDEVATCLRRKRYSEILNVKIASPQFSTRFGPIVDGLVIPNTPHKVMGQYSDIFSGYD
Cqui            -DNDEVAACLRRKRYSEILNVKIASPQFSTRFGPIVDGLVIPNTPHKVMGQYSDIFSGYD
Agam            -ENDEVTACLRRKRYSEILGVKTASPQFSTRFGPIVDGLVIPNMPHKVMGQY-------D
Dmel            EDDDVLLDCLRHRRYQDILHIPTALTQFSTSLGPIVDGHVIPNQPYKVMGHYTEHFSRYD
Dana            EDDDVLLDCLRHRRYQDILHIPTALTQFSTSLGPIVDGHVIPNQPYKVMGHYTEHFSRYD
Dpso            DDEDALLDCLRQRRYQDILHIPTSLQQFSTSLGPIVDGHVIPNQPYKVMGHYTEHFSRYD
Dgri            ---DALLDCLRQHRYQDILHIPTSLQPFSTSLGPIVDGHVIPNQPYKVMGHYTEHFSRYD
Dvir            ---DALLDCLRQRRYQDILHIPTSFQQFSTSLGPIVDGHVIPNQPYKVMGHYTEHFSRYD
Dwil            ---DALLECLRQRRYQDILHIPTTLPQFSTSLGPIVDGHVIPNQPYKVMGQYAEHFSRYD
Dpul            --DDELAACLRFKRVTELMSAPLYAS----PLAPLVDGIVVPNEPRQSMKIYNELFGRYD
                   * :  *** :*  :::           :.*::** *:** *   *          *

Llon            LLFGMTEVESYHMLNAIELMYGLLENERDTLLRFYLQNRFEIRPDLALAATLKEYSNIYA
Ppap            LLFGMTEVESYHMLNAIELMYGLLENERDTLLRFYLQNRFEIH-----------------
Aaeg            LLYGMTELESYNILNAVALTYGLLENERDNLLRFYMQNRFEIRPDLALAATLNEYTGILT
Cqui            LLYGMTELESYNVLNAVALTYGLLENERDNLLRFYMQNRFEIRPDLALAATLNEYTGILT
Agam            LLYGMTELESYHILNAVALTYGLLENERDNLLRFYMQNRFEIRPDLALAATLREYTDIYM
Dmel            LLFGITESESYHTLAALALEEGLRENERDNLLRFYMQSRFDIRPDLALAATLKKYQDMYN
Dana            LLFGITESESYHTLAALALAEGLRENERDNLLRFYMQSRFDVRPDLALAATLKKYQDMYN
Dpso            LLFGITESESYHTLAALALEEGLRENERDNLLRFYMQSRFDVRPDLALAATLKKYQDMYN
Dgri            LLFGITESESYHTLAALALEEGLRENERDNLLHFYMQSRFDVRPDLALAATLKKYQDMYN
Dvir            LLFGITESESYHTLAALALEEGLRENERDNLLRFYMQSRFDVRPDLALAATLKKYQDMYN
Dwil            LLFGITESESYHTLAALALEEGLRENERDNLLRFYMQSRFDVRPDLALAATLKKYQDMYN
Dpul            VMYGVTQSESFHLLNAATLQFGLTETERNRIIRTYVRNSYASNIDQVTAAILNEYTDWKN
                :::*:*: **:: * *  *  ** *.**: ::: *::. :  .

Llon            TEKTQNANFHRDKILEIFSDARVAGPLVQTGLFHAKSNPRSFMYVFAHNSEAGEYAYLSQ
Ppap            ------------KVLEIFSDARVAGPLVQTGLFHAKSNPRSFMYVFAHNSEAGEYAYS--
Aaeg            DPNKSLADIHRDTLLDILSDARVAAPMLQTGLYLAKVNPKCYMYVFGHNSEAGEYGRLSQ
Cqui            DPNKSLADIHRDTLLDILSDARVAAPMLQTGLYLAKVNPKCYMYVFAHNSEAGEYGRLSQ
Agam            DPNKALADEHRDNLLEILSDARVAAPMVQTGLYLSKVNPKCYMYVFGHNSEAGEYGRLSQ
Dmel            NPIKATNLEHRDVVLDILSDARVVGPLLQTGMFHADVNRRNYMYVFGHNSATGPFAHLPH
Dana            NPIKATNLEHRDVVLDILSDARVVGPLLQTGTFHADVNRRNYMYVFGHNSATGPYAHLPH
Dpso            NPIKATNLEHRDVVLDILSDARVVGPLLQTGMFHADVNRRNYMYVFGHNSATGPYANLPH
Dgri            NPIKATNLEHCDVVLDILSDASVVGPLLQTGMFHADVNRRNYMYVFGHNSAMGPYANLPH
Dvir            NPIKATNLEHRDVVLDILSDARVVGPLLQTGMFHADVNRRNYMYVFGHNSATGPYAHLPH
Dwil            NPIKATNLEHRDVVLDILSDARVVGPLLQTGMFHADVNRRNYMYVFGHNSAMGAYANLPH
Dpul            --PVRDTEEYRDSILEILSDARVVAPVIQMADLHSSIRQRSYFYVFAHQTTNGDYPQNWG
                            :*:*:*** *..*::* .   :. . : ::***.*::  * :

Llon            SIIGEELAYVFGAPLSPSGPFQP-FYNSQEKLLSEAVMKYWTNFAKTGNPKAPWRDKFLN
Ppap            -IIGEELAYVFGAPLSSSGPFQP-FYNSQERLLSEAVMKYWTNFAKTGNPKAPWRDKFLN
Aaeg            SVVGEDLAYIFGAPLGPVGPFQT-HYNARERLFSEAVMKYFSNFAQTGNPKAPWKDLFLN
Cqui            SVVGEDLAYVFGAPLGPVGPFQT-HYNARERLFSEAVMKYFSNFAQTGNPKAPWKDLFLN
Agam            SVVGEDLAYVFGAPLGQVGPFQH-HYNARERLFSEAVMKYFSNFAKTGNPKAPWKDLFLN
Dmel            SIMGEELAFIFGAPLAAAGPFPSGNYTVQEKLLSEAVMAYWTNFVKTGNPKAPWKGIFIN
Dana            SIMGEELAFIFGAPLAPAGPFPSNNYTVQEKLLSEAVMAYWTNFAKTGNPKAPWKGTFLN
Dpso            SIMGEELAFIFGAPLAPAGPFPSHNYTMQEKLLSEAVMAYWTNFVKTGNPKAPWKGIFLN
Dgri            SIMGEELPFVFGAPLAPVGPYPSHNYSVQEKLLSEAVMAYWTNFVKTGNPKAPWKGTFLN
Dvir            SIMGEELAFVFGAPLAPAGPFSSHNYSVQEKLLSEAVMAYWTNFVKTGNPKAPWKGTFLN
Dwil            SIMGEELAFIFGAPLAPAGPFPSHNYSVQEKLLSEAVMAYWTNFVKTGNPKAPWKGTFLN
Dpul            SIHGEELAYVFGMPLVGGTNHLSSNYTRAEMLLSEIVITYWANFARTGNPNFPPRQKFLT
                 : **:*.::** **     .    *.  * *:** *: *::**.:****: * :  *:.

Llon            LNP--IEWARFDIDWPEYNQINQSYLHLNSHPVVGHQYRHTFMRFWNQNLPSELHKLTTA
Ppap            GNP--IEWARFDIDWPEYNQINQSYLHLNSHPAIGHQYRHIFIKFWNQNLPSELHKLTTS
Aaeg            MNP--EDWRYYDVDWPEFNSVNQSFLHMGLTPVVSRHYRQQYMKFWNQKLPEELKRITTS
Cqui            LNP--EDWRYFDVDWPEFNSVNQSFLHLGLTPVVSRHYRQQYMKFWNQKLPEELKRITTS
Agam            LNP--EDWSYYDVDWPEYNSINQSYLHMGITPVVGHRYRQKYMKFWNEELPEELKKITSS
Dmel            SHA--LEWDRYDLDWPEFNRRAQAYLNIGIPPTVGYKYRQIYMNFWNKELPDELNQIAAI
Dana            SHA--LEWDRYDLDWPEFNRRSQAYLNIGIPPTVGFKYRQIYMNFWNKELPDEINQIAAI
Dpso            SHA--LEWDRYDLDWPEFNKRAQAYLNIGIPPTVGYKYRQIYMNFWNKELPDELNQIAAL
Dgri            SHL--LEWDRYDLDWPEFNKRAQAYLNIGIPPTVGYKYRQIYMNFWNKELPDELNQIAAI
Dvir            SHL--LEWDRYDLDWPEFNKRAQAYLNIGIPPTVGYKYRQIYMNFWNKELPDELNQIAAI
Dwil            SHA--LEWDRYDLDWPEFNKRAQAYLNIGIPPSVGYKYRQIYMNFWNKELPDELNQIAAI
Dpul            VANSRDRFEPSYVQWPAYDRHSQKFLNIDLRPKVKDHYRADKMALWSKLIPELMSNGNGK
                       :    ::** ::   * :*::.  * :  :**   : :*.: :*. : .

Llon            RTNIHMMPPSIPSAEGPRRE--------IYGNFSTHPRHFGGG-EVRSRPDDMMYAKTEN
Ppap            RSSVHMIPPNIPSPDVPRRE--------IYGNFSTHTKHFGGGGEIKTHPDDRMYSKTDN
Aaeg            KPYSPYSELISPP----------------DRRMTTSSPEFVTGHINLFPIKVDIEKPTED
```

```
Cqui            KPYSPYSEFISPP-----------------DRRMTTSNPEFVTGHINLFPIKVDIERPTED
Agam            KAYAPYSDFFSPPGAGGKQGGGVSAVVGRMRGGTTPHPDYPTGHINLYPIHVDIERPTED
Dmel            QEQ------------------------------LQKPGQEVITGHMSKYGPR---DHGAED
Dana            QEQ------------------------------LHGPSHEVITGHMSKYGPR---DHDSED
Dpso            QEQ------------------------------LHP-GSEVITGHMSQYGPR---DKGAED
Dgri            QQK--GQQQQQQQ-------------------QQQFPGNGVITGHMSKYGAR---DKGAED
Dvir            QVR--GQQQQLN-----------------------GNEVITGHMSKYGAR---DNGAED
Dwil            QQQQHHQQQQQQE-------------------SNVVAGSNVITGHMSNYGSR---DNGAGD
Dpul            VFTVGNEDDDDDD--------------------DGEVLVTGGGEASSVAEPDSSSHSPPA
                                                          *             .

Llon            PIAILRLMNKNTRPSHGVTVSEEKMSTEAANVAPFD-------------------------
Ppap            PMTILRLLNKSPRPTPEIHSFEDKVSTEVGDVSPFD-------------------------
Aaeg            SFKALVHRMKDPGMYNAPASSIGIPAEAPVKEIPKP-------------------------
Cqui            PFKALVYKMKDAQVGAPAPAAVGMPAEAPAKEAPKP-------------------------
Agam            PFRELLYQMKDPLAGPGMYNAPATSSAAAIPEKGASPAQLAAGVLEHPKH-----------
Dmel            PVRTLKLLLQEPSLAGPTQSGESTETAAENMYNAPPTFGHVHKMQGGSDF-----------
Dana            PVRVLKQLLQEPTLAGPE----STETAAENMYNAPPTFGHIHKIQGGSDF-----------
Dpso            PVRTLKLLLQEPSLANSP---DSTETAAENMYNAPPTFGHVMHKQQPQQP-----------
Dgri            PVRTLKLLLQKPLPSKGS---EQAETAAENMYNAPPTFGHVHRQQQQQQP---------A
Dvir            PVRTLKLLLQEPLAT-GS---EQLETAAENMYNAPPTFGHVHRQQQQQQQQQQQGIGTNG
Dwil            PVRTLKMLMGVVGEP--------TEAAAENMYNAPPTFGHVHKQQGEGGE-----------
Dpul            VMGAQPEVYQQPSLGGFEVFPIGPMSKGASGNSSRP-------------------------
                 .                      :

Llon            -------------------------------PDSYVPNSANLTLNLLIIIVVILSILIVIF
Ppap            -------------------------------PESYIPNSANLTLNLLIIIVVILSILIIIF
Aaeg            ----------------------ENVEKDMENSEMMKTEASLTLLITVAIVFLIFNIFVIV
Cqui            ----------------------ANVEQESENSEMMKTEASLTLLIAVAIVFLLFNIFVIV
Agam            ------PKQLQQSGKDGPFGVTDEQELDGGATEIMKSESTLTILIAVAIVFLLFNIVVIV
Dmel            ----------EDLVTSTNSLESGEDHPQAPPETVAKSEATMQLLIALITIIIVLNLLIYG
Dana            ----------ELITPPSSFDNGDDQQRSESPETVAKSEITLQLLIALIAIIIALNVIIYG
Dpso            ----------QREE-DLEMTSPTNPMAEDQPETVAKSETTLQLLIALIALIIALNVIIYG
Dgri            VEGVAISGEDPTQTSQHVSDYGEYGATVAPTEPIAKSETTLQFLIALIAIFIVLNVVIYA
Dvir            AEGVAISGEDTTHSNDYG-EYGASSSTATPAEPIAKSETTLQLLIALIAIFIVLNVAIYA
Dwil            --------EMVTGIPFSPGSNDVKTESSSSTETVAKSEITMQLLIALIAMIIILNVIIYG
Dpul            --------------------------DDGQDLVPQTGGIALSIVIVIGICFLVLNVCACA
                                              : : : :   .:   :

Llon            GYILKRSYKR--------------SSPTEKILKNLKRSTPPKMFYRSDNDLSEMSNN---
Ppap            GYILKRSYKR--------------SSPTEKMLKNLKRSTPPKMFYRSDNDLSEMSNN---
Aaeg            GYIIKRNVGN--------------RKMKRKFDDNILESAASMNEKRSKLNDTEDSY----
Cqui            GYVIKRNVGN--------------KKVKRKFDDNILESAASMNEKRSKLNETEESY----
Agam            GYLVRRHLGRPGVGGVGVGGVAGGRKVKRKYDDTMLESAAAVVVSGVGLGEKEDGCNQLH
Dmel            TFLLRQRRRR----------AKALPFPAPKLGGTILSYDGANDEELKRCSKSRDGDDS--
Dana            TFWLRQRRRR----------SKALPFPGPKLGGHILSYDGANDEELKRCSKSRDGDDS--
Dpso            TFLMRQRRKR----------AKMLPFPGPKLNGNILSYDGANDDELKRCSKSRDGDDS--
Dgri            TFLFRQRRKQ----------NTSLQR---KLGGQMLSYDGANDDELKRCSKSRDGDES--
Dvir            TFLLRQRRKR----------NTSLQR---KLGGNILTYDGANDDELKRCSKSRDGDDS--
Dwil            TFLIRRRRAK----------TSPLKR---KLGGNILSYDGANDDELKRCSKSRDGDDS--
Dpul            GVFYQRDRVR----------------FKEMLIQRQYKLKSANHDGEGPTSGAQVAAAG--
                    ::   .                          .       .  . .

Llon            ------------------SYDMVRIERGSKVDSIHSDG----INYPTTVIKAPWRTLPKR
Ppap            ------------------SYDMVRIERGSKIDSIHSDG----INYPTTVIKAPWRTLPKR
Aaeg            ------------------ILDLMRKSNPYEPVAKANN-----FTLPRQFSGETMDPHTKV
Cqui            ------------------ILDMMRKTNPYEAVVKGNN-----FTLQRQFSEDTVDSHTKV
Agam            GHHHHHHLPHGAYLLDEAATMMLRKSNSYEPVAKQPFGG--ELHGPVSIGVDAVDAHTKV
Dmel            -----------------FVLEMTRKSNTYEAIKTGQRS----------LSCSTVDTHTKV
Dana            -----------------FVLEVTRKSNTYEAIKTGQRS----------LSCSTVDTHTKV
Dpso            -----------------YVMDMARKSNTYEAIKTGQRSPINGYAIQRQLSCSTVDTHTKV
Dgri            -----------------YVLDIVRKSNTYEAIKTGQRSPINGFAMQRQLSSSTVDTHTKV
Dvir            -----------------FTLDMVRKSNTYEAIKTGQRSPINGYAMHRQLSSSTVDTHTKV
Dwil            -----------------YILDVVRKSNTYEAIKTSQR----------QLSCSTVDTHTKV
Dpul            --------------ARETLLALHRVEEDFPELDGSANG--MQSGLPHQASTSTMDPHTKV
                             : *  .                          .  . .*

Llon            CSFEDPDNQ-------------------SDCFIKPGQVSVAIDATPQARSGSVLKQEPIE
Ppap            CSFEDPDNQ-------------------SDCFMKPGQVSVAIDATPQARSGSVLKQEPIE
Aaeg            CDWISQDYGTVCTKQIECDGNG---SCNKDNLPLPQKVSVAIDATPQARSNSVLRQEPIE
Cqui            CDWIAQDYGAGCPNQTQCEGLGSIPSCNAATLNVPHKISVAIDATPQARSNSVLRQEPIE
```

```
Agam            CDWMVASAG-----------------CRMAPAVPPHKISVAIDATPQARGNSVLRQEPIE
Dmel            CEWMSSQ---------------------EAPKSGSFNTATPPP-QPILSDGRLIICQDIE
Dana            CEWMSGQ---------------------EAPKAGSFATATPPPHQPILADGRLIICQDIE
Dpso            CEWMSGQ---------------------EVPKSGSFATPTATTPTSVLPDGRMIICQDIE
Dgri            CEWMSSQQQSP---------------TQKLLKSGSFNTTPAPLPAAQPD-NRIIICEDIE
Dvir            CEWMSAQHQHTN-------------SGHKLPKSGSFNTTPPPPPAVQSD-SRIIICQDIE
Dwil            CEWMSGHPSSHN-------------HTLPKSGSFATNTPSPPPAMANADGSRIIICQDME
Dpul            SQWMAQEITIER-----------------CPTPPINNRLAKPAVQDKLGGSDRYDSRGYD
                ..:                                              .     .  :

Llon            VTKSRILIAPETKIICRELDNPPLTVYYGDDLCTEEEYPRKPEENHYNTMGNGHVRNFSD
Ppap            VTKSRVMLAPETKIICQELDNPRISAYYGNDLCDEDDYVRKPPDSGFSTLGNGHVRNFSD
Aaeg            ITKAKSFECASPEDYSESVS-------QDSPCDSDIRQQCSSSGSSYSYNDCDYKVEPKH
Cqui            ITKAKSFEYASPNDYDESVS-------QDSPTDSDMRQQCSSSGSSYSYNDCDYKIEPLL
Agam            ITKAKSFEYGATDEVDCAVSGCPGEGRMGEGNASGTGSSSSGTSSSCSSSATTDELKQTM
Dmel            VADAALLIPQHMHEPQH---YEMLLQRQHSALTEPSDEILQQQYPLR---NHSHSHSDPV
Dana            VADAALLIPQHMHEPQH---YEMLLQRQHSALTEPSEDVLQPQYPTR---SHAHSHSDPV
Dpso            VADAALLIPQHMHEPQH---YEMLLQRQHSALTEPSEDILQQQLQHMP--LHRHTHSDPV
Dgri            VADAALLTPQHMHEMQEGLHYELLLQRQHSALTEPSEATLQPQH--------SHTHSDPV
Dvir            VADAALLTPQHMHETLEGMHYELLLQRQHSALTEPSEATLQPQHGHTYAH--AHSHSDPV
Dwil            VSDLALLTPQHMHDMQH---YELLFQRQNSALTEPSEDVLQPATHHQHHHHHQHSHSDPV
Dpul            MEDPASALFPLLTKSPNHQDIYGLLPVKEENSADAKQPITSGQSPDFINQVHNNGTFLKF
                : .                          .         .

Llon            SQKGGEVSAVIVRGTNS-----------LDSMKMRRRSLPAQNQLDSALGWTKVPPMPPP
Ppap            TQKCGELPSTFARKTNS-----------LDSMKMRRRSLPAQNQLDFAMGWTKVPPVPPP
Aaeg            FKSDSNPIYNNYPYQN-----------EEITSFIEP----GDINVTSRDKSEEKDPLSPE
Cqui            SKSDANPVYNNYPYQTT----------EDVTSFIEP----GDINVTSREKSQEKDPLSPE
Agam            PGAGAGLYYNCPDYPVRRFGATGEGGEEEVTSFIEQDTPAGDINVTSRDKSTEKDPLSPE
Dmel            DMILAA---------------------DEQVTSFVHADDVDINVTSRDDSDGLEVIPLTSA
Dana            DMILAA---------------------DEQVTSFVHADDVDINVTSRDDPGVLEIIPLTAA
Dpso            DMILAA---------------------DEQVTSFVHADDVDINVTSRDENCVLEVVPLTAA
Dgri            DMILAA---------------------DEQVTSFVHADDVDINVTSRDE-CELQLQPLTVA
Dvir            DMILAA---------------------DEQITSFVHADDVDINVTSRDE-DALVLQPLSAA
Dwil            DMILAA---------------------DEQVTSFVHADDVDINVTSRDDRLDNDAVPLTAA
Dpul            GDVGHVG--------------------ESSPPFADSASSAQTMSRSSTGRRKARSKSQLG
                                               .

Llon            RTVSTLSRRHNPPVPEENLPHP-----------------------------------------
Ppap            RTVSTLTRRQNPTVPEEILPHP-----------------------------------------
Aaeg            EALKVIQRRNFPKVLPDYPNG--------------VAGITASMKRRSLPPQAFLLANSHS
Cqui            EALKFAEGRQSPHSGTTTPN-------------------------------------------
Agam            EALRVIQRRNFPKVLPDHPRTGGTGGPTTVGTSMKRRSLPPQSLYGALPHGNCHSLRREA
Dmel            QQLELLRQRNYPKVLPTEQDLINSS----------------------------YKRNSLP
Dana            QQLELLRQRNYPKVLPTEQDFLDTS----------------------------YKRNSLP
Dpso            QQLELLRQRNYPKVLPTAQDLRDSRG---------------------------GNYKRNSLP
Dgri            QQLALLQQRNYPKVLPTEQDVR-------------------------------QSYKRNSLP
Dvir            QQLELLRQRNYPKVLPTAQDLRTS-----------------------------HSYKRNSLP
Dwil            QQLELLRQRNYPKVLPTAHDLAAKGVL-------------------------AVNFKRNSLP
Dpul            SQRSITTKRDVAVGGDGDDDEGNYRVAP---------------------------DNGGED
                       *. .

Llon            ---------------------------------PQHMLATLPPSSDLEDEVITSNTLIVG
Ppap            ---------------------------------PQHMLATLPPSSNLEDEVITSNTLIVG
Aaeg            LRRDSSRLVPAPPPRISSTLGRRSSKNRTSNNFLSSPPVMAEEPPIEEEPPITLNTLHVG
Cqui            ---------------------------------LIDPGKEIFQAPIQQQFP----------
Agam            AGAGRLIPAPPPRICSTLGRPAAVRQARASNNFISSPPIVAEEPPVEEEPPIALNTLHVG
Dmel            PQNFNAPLPPPRTISNTLGRRRRDSSNITTSPLQVARDCGGEDED-LKEPQITQNTLIVG
Dana            PQNYNAPQPPPRTISSTLGRRRRDSSNVTTSPLQLAQDISADDEDNLREPQITQNTLIVG
Dpso            PQNFSAPLPPPRTISSTLGRRRRDSSNITTSPLQLAQDLAGEEEP--GEPQITENTLIFG
Dgri            PQNFQAPLPPPRTISSTLGRRRRDSSNITTSPLMLAHDNGEDELH---EPQITQNTLIVG
Dvir            PQNFTAPLPPPRTISSTLGRRRRDSSNITTSPLMLAQDSGEEEPR---EPQITQNTLIVG
Dwil            PQNFNPPVPPPRSTSQTLGRHRRDSTNVTTSPLQLARDCGEDEPG---EPLITQNTLIFG
Dpul            DRRASALYGYGAMDTTIRRLNLPKVLPDLPHQDVPAEATAPSGVDLTGPKTSSSNLVPSD


Llon            PLIPTKGTKGKSSDASDD----------------------------------------
Ppap            PLIPSKGPKGKSSDASDE----------------------------------------
Aaeg            PLLPKHHENTYMTMSRQN----------------------------------------
Cqui            ----------------------------------------------------------
Agam            PLLPTHQESTYMTMSRQN----------------------------------------
```

```
Dmel            PIVPKSPASSLKRVKR-------MPESSAMTALSGSFQSFEAVPPAHETTPPQGGERTEC
Dana            PIVPKSPATSLKRRKDK------EPAAAAMTALSGSFQSFESVPPAHETTPPQG-ERTEC
Dpso            PIVPKSPATSLKRQKRTP--EAGAPAAGAMTALSGSFQSFESVPPAHETTPPQGGERTEC
Dgri            PIVPKSPAASLKR-------IKRQTDTAAISKLSGSFQSFESVPP-AHEATPPQGQRTEC
Dvir            PIVPKSPAASLKR-------IKHTGVATALSTLGGSFQSFESVPP-AHEATPPQGQRTEC
Dwil            PIVPKSPATSLKRPKPRTLPGSETELPTAMNALSGSFQSFESVPPPVHEDTPPQGERTEC
Dpul            QHSPIYTIPYAGLPTGVDSGERSTPISRNERHPRQTMESPPASTVPLSADTKKQLSKSRR

Llon            --GDQDEATLSG---------VASTTKYNALSRSKKKSTASKNEPNI-------------
Ppap            --GESAEATLSG---------TSSTTKYNALSRSKKKSTS-RNEPNI-------------
Aaeg            --SAETEQTT----------STEEPAVDIICIEHKPESHYSYIKPPS-------------
Cqui            ------------------------------------------------------------
Agam            --SLGSECEPGSSEADDDRPSEEQPPVDIICIEHKPENHYSYVRPVG-------------
Dmel            IYAIRPS---------PGSCSWAAPNGDLYAQPMKSSSRNSLIPRPTKPAESQS------
Dana            IYAIQASPASPTTPVSPGSSSWSAPNGDLYAQPVKSPSRTSLIPRPTHSLQVAT------
Dpso            IYSIEASPG---TCPGNDGTSWSAPNGDLYAQPMKSPSRTSLIPRLTQQQPTPTPTPTPT
Dgri            VYAIEASASVSSLP------SWSAPNGDVYAQPTKSASRTSLIPRLLQTAVETQQVST--
Dvir            VYAIAASASVSSVPGISSPSSWSAPNGDLYAQPTKSASRTSLIPRPLQTAPAAAAAEAP-
Dwil            VYSIEPTGSGPSSPG-----AWSAPNGDLYASPIKSPSRTSLIPRPMQTAAAPAKTTTE-
Dpul            KYAGSAGGPANTSGGSSEAVLLSSPTSTIVKTNLPQTVVVAPHPRANRVAASSPTAGLLH

Llon            ----------------------------------------IIKPQITRSTSEKANKFNARR
Ppap            ----------------------------------------IIKPQISRSTSEKANKFSARR
Aaeg            ----------------------------------------TIKPPTSFKSSPPKELNRTES
Cqui            ------------------------------------------------------------
Agam            ----------------------------------------LMRTPSSFKCPEAVSRTEGGE
Dmel            -------------------QATPAGPAGESLGSTGMATASRIPQLQRQASGKDLQAR---
Dana            --------------------SQAPQTGDGVPG-----AIPSRIPQLHRQASGKDLQATLQA
Dpso            PTPTPTPSPHHHHQGAAEGEETSAVPGGASSMATTVSPTRIPQLLQRQASGKDLRIANTA
Dgri            ----------------------EATAKAT---TAAAAAMPSRIPQLQRQASSKELQSPRES
Dvir            ----------------------AATQHVS---TAADSATPSRIPQLQRQASGKELSPVDGQ
Dwil            ----------------------PQSPTCVGDAEAVLNPNPTRIPQLQRQASGKDLYQPTLR
Dpul            Q------------------------------QQQPEPCLVIRPGPRQTCQSTSTASTAGG

Llon            MPAAANSAGTTETPD-------------------------------------------------
Ppap            IPASSTTTNPAEANETNI----------------------------------------------
Aaeg            RAELQQASTYGSTPRLPITRSGSSIMHREMSSDSSATDTTSGSTGTIKML----------
Cqui            ------------------------------------------------------------
Agam            ARSAGDGETGQPTP-IDAATQPLQPVRDKFSSDSSATDTTSGSTGTIKK-----------
Dmel            TATDNTANPLGCQPRTDSTISSGSSAS----YASTTDSSSSSSTGTVRTELQQFQPAP--
Dana            AFTENTANPSGCQQRTDSTISSGSSASNDSSTSSSTGSSSSSSTGTVRTELQQFQPLP--
Dpso            SPLASQGQQQQQQQQQRTDSTISSGSSASAASSTESSTSSSSSSTNTVRTELQQHQHQHQP
Dgri            LALLDSTDGQSSNASADSTSRTKSSISNDSSWRSSSASSSSSSTGTVRTELQQYQQPSKT
Dvir            QTVDATTPTSNSRVGSTISS---DSSASAASWRNSTTSSSSSSTGTVRTELQHCQQSGRT
Dwil            TNSTISSGSSASAASGTSAT--------------SATSSSSSSTGTVRTELQKSPSQQQQ
Dpul            PAGVSGYDSATNDNEFYPSATADVILRRPRQQDASTNNPTRPANRNSRSWYAQYSQSFIS

Llon            ------------
Ppap            ------------
Aaeg            ------------
Cqui            ------------
Agam            ------------
Dmel            GRSITTNI----
Dana            GRSITTNI----
Dpso            GKSITTNI----
Dgri            IATQI-------
Dvir            IATHI-------
Dwil            SSKCITTKI---
Dpul            QSIDQESDKNDN
```

Figure 5. Neuroligin 1 Homologue Alignment. The yellow, highlighted regions indicate the transmembrane domain for each protein sequence.

Figure 6. Neuroligin 1 Phylogenetic Tree

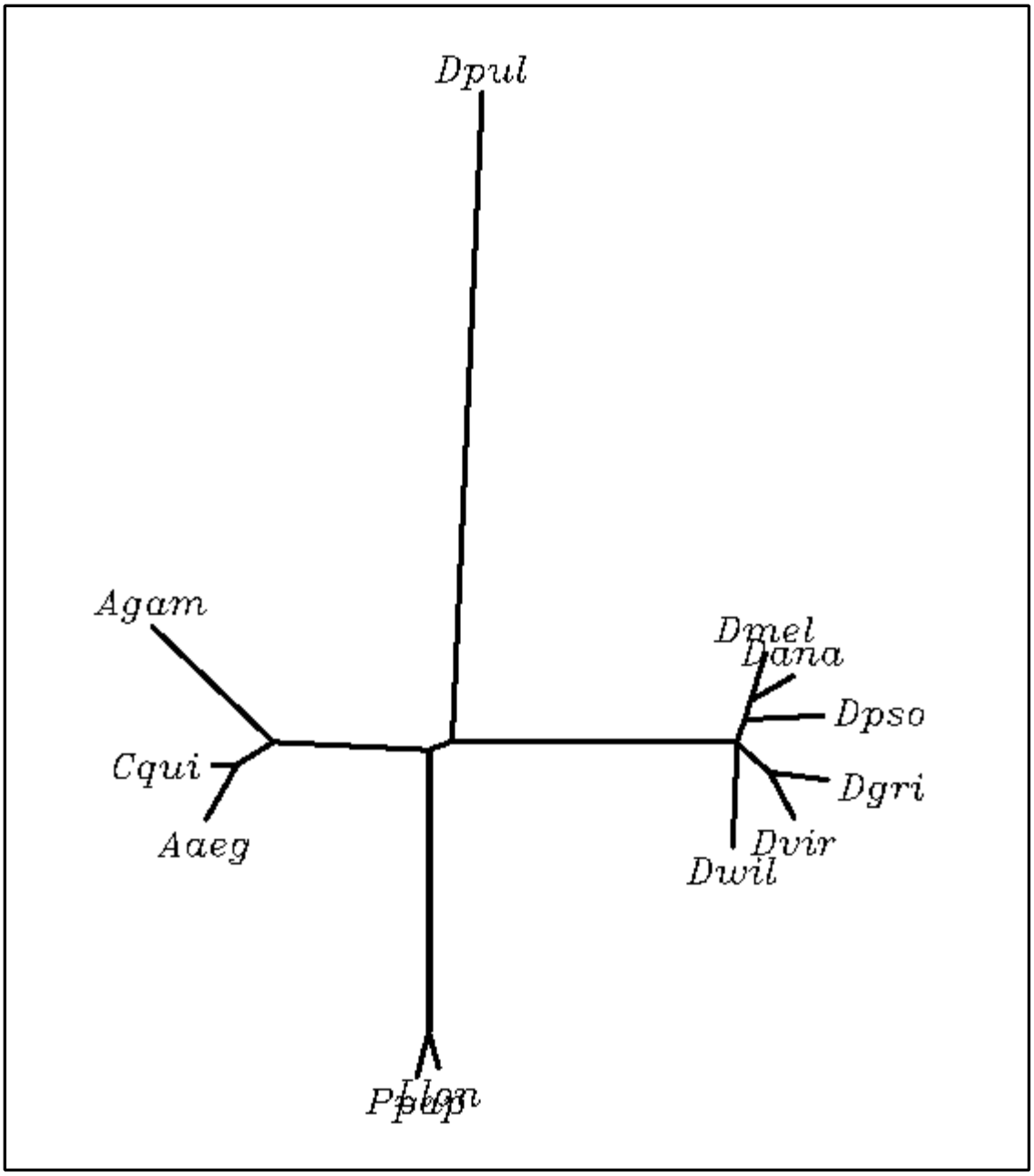


Table 3. Neuroligin 2 Homologues

D. melanogaster	MAQTKIIQSRFLHIYGLLCLGSLMGCIKTIGASIAQQTLADADAPDEAKAKEREAE TDADADAEGDAKAETETGKGTREHTNGDAEADARIMATGICIIRLITLKRFLENSK TEEHNQRRQQSATSAASAAPAGATSPRRRHLHYPGHSAVQPGHPEACALLVLLLL TSLWPDCCECLHGGSNTVKTKYGLLRGIVVRSSPLVEAFLGIPYASPPVGSLRFMP PITPSTWKTVRSADRFSPVCPQNIPIPPNGPEALLEVPRARLAQLRRLLPLLKNQSE DCLYLNIYVPYETRRQRRNTDDTTGEPKTKLSTVVFIHGESYDWNSGNPYDGSEL AAHGNVIVVTINFRLGIFGFLKTGGKESAQGNFGLMDLVAGLHWLKENLPAFGG DPQSITLLGYGTGAVLANILVVSPVASDLIQRTVLVSGSALSPWAIQKNPLFVKRR VAEQTGCHGDMLYDDLAPCLRTKSVAELLAVKVDHPRFLVGFAPFVDGTVISPG ANPLGSTTLPLGSAIVSTSGIEYANFPKRDLIFCLTSVESYLDLSAQDLEFGFNETRR DRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRNPLSSRDATLQFLSDGHTASP LIKLGYMHSLRGGRAYFLHFKHKTIEEEYPQRSGSVRGEDVPFWLGLPMSPLFPH NYTTQERQIGRLMLRYLSNFAKTGNPNQSTAKSVLPNPNEVLETALHQQKKRSTS LTHPNLSEALNLAVLYNQRRSNAMHEKRSYIRRRLRSNDAAFTQLGISSERDVGS YDGDELPFWDAYDVVNQLYVELGNKANIQSHYRGHKLSMWLNLIPQLHRHFNIN DQSMRHHQFQDDMNNRDLYEGVVRPQLQTKPAEDDNIIIMQRSRTTTPPPPSKNP STNATQALNPLGTGTPTTTECGIDGAMSVSELTTTQSPKDNRTGIVRVETQKDLAT ASTGIIGNLELLRRLGGKQFQSYTTALIATVAVGCFLLILNVLIFAGIYHQREKRAR DAKTKEELQEGDNSKNSSILKLNALMGGGGACGSGPGDIADAYTLSGSVVDSKG GVGVVFGEYSCYDEKTKQLKEEKLLVELPPSSSTGQTSLMIDNWGACSTSTLDLL KTKPGDPIEMVTYSALPSVMESTSAMSSKRGSFVDTTLQFNSPQQFDYAVQSSDQ MSFKAIEDAVKAAAGHDSEITRDDDIPEPPPPPRSFLAAQQQQQQQQQQQQTGILR QTGAGGTSTSGSASGSGKKRVHIQEISV
D. ananassae	MTTLHMNQKLNMRTTAVKMKRTRYPASCPSRSQTVAEADGADKEREAEPETGA GARAEAGSREGAEAKRRAETEADAEADARIMATGICIIRLITLKRFLEISTTNEHNQ RRQQSATSAGVTSPRSRRRHPHSHPGQPGHTASHHTAGHPEACWRHQLTPSTLL MLLLLTSLWPDCCDCLHGSSNTVKTKYGLLRGIVVRSSPLVEAFLGIPYASPPVGS LRFMPPITPSTWKTVRSADRFSPVCPQNIPIPPNGPEALLEVPRARLAQLRRLLPLL KNQSEDCLYLNIYVPYETRRQRRNIDDTNGEPKTKLSTVVFIHGESYDWNSGNPY DGSELAAHGNVIVVTINFRLGIFGFLKTGGKESAQGNFGLMDLVAGLHWLKENLP AFGGDPQSITLLGYGTGAVLANILVVSPVASDLIQRTVLVSGSALSPWAIQKNPLF VKRRVAEQTGCHGDMLYDDLAPCLRTKSVAELLAVKVDHPRFLVGFAPFVDGT VISPGANPLGSTTLPLGSAIVSTSGIEYANFPKRDLIFCLTSVESYLDLSAQDLEFGF NETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRNPLSSRDATLQFLSDG HTASPLIKLGYMHSLRGGRTYFLHFKHKTIEEEYPQRTGSVRGEDVPFWLGLPVSP LFPHNYTTQERQIGRLMLRYLSNFAKTGNPNQSTARPTLTIPSEESDVSLHQQKKR STGLIHPNLSEALNLAVIYNQRRSNAMHERRSYIRRRLRSNDASFTQLGISSEKDV GSYDGDELPFWDAYDVVNQLYVELGNKANIQSHYRGHKLSMWLNLIPQLHRHF NINDQSMRHHQFQDDMNNRDLYEGVVRPQLQTKPAEDEHIIITQRSRTSTTAPPPP KDATPNATQVLNPSATGATTTTTDCNIDGAMSVSELTTTQSPKDNQTGRVRVETQ KDLATASTGIIGNLELLRRLGGKQFQSYTTALIATVAVGCFLLILNVLIFAGIYHQR EKRARDAKTKEELQESDTSKNSSMLKLNALISGGGGGGHGAGDMADAYTLSGSI VDSKGGVGVVFGEYSCYDEKTKQLKEEKLLVELPPASSSGQTSLMMDSWGACSS STLDLLKNKPGDPIELVTYNTLPTVMESTSALSSKRGSFVDATLQFNCPQQFDYAV QSSDQMSFKAIEEAVKAAAVNDSEITRDDDIPEPPPPPRSFIAAQQQQQQQQQTGIL RQTGAAGSSTTGSGNTSGKKRVHIQEISV
D. grimshawi	MTTTNLRSKKQQPTKLKLNQSRFLRIYCYVLVGSLVSYIRAICATTVKKLQMTAE TEMEIRSISVSNSNSDSNSDSSLQLRTRICIIRLITLKRFLHILAMPTHNQQQQQQQQ QHHQHQHHQQQQQLQSAHRKRHLWHSQVLTQAPSCWRHQTSGSTWLLLLLLLS SLWPDCECVHGAKLSANTVKTKYGLLRGIVVRSTPLVEAFLGIPYASPPVGSLRF MPPITPSTWKTTRNADRFSPVCPQSVPIPPNGPEALLEVPRARLAQLRRLLPLLKNQ

	SEDCLYLNIYVPEEARRTQPDNEDADSSSSSSSSDSLLSTVVFIHGESYEWNSGNP YDGSELAAHGNVIVVTINFRLGIFGFLKTGGKESAQGNFGLMDLVAGLHWLKEN LPAFGGNPQSITLLGYGTGAVLANILAVSPVASDLIQRAVLISGTALSSWGIQKNPL FVKRRVAEQTGCHGDMLYDDLAPCLRTKSVAELLAVQIDHPRFLVGFAPFVDGT VISPNTETMSSTALPLGSAIVSTSGIEYANFPKRDLIFCLTSVESYLDLSAQDLEFGF NETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRNPLSARDATLQFLSD GHTAAPLIKLGYMHSLRGGRTYFVHFKHRTIEEEYPQRTGSVRGEDVPFWLGLPV SPLFPHNYTTQERQIGRLMLRYLANFAKTGNPNQSTSSTALAAAATSQTQVATHT QRKRSAPRYLDAMATGEALNLAVIYNQRRSNAMHEKRSYFRRRLRSNDGVSTQQ GLGTELEPGGYDGDALPFWDAYDVVNQLYIELGNKVNIQSHYRGHKLSMWLNLI PQLHRHFNINDQSMRHHQFQDDLNNRDLYEGVVRPQLQTKPAEDDNILVIQTTRT TTTPLPPQSPKEENTNATNALSATGATATTECGIDGAMSVSELTTTQSPKDNRTHS KVETRKDLATASTGIIGNLELLRRLSGKQFQSYTTALIATVAVGCFLLILNVLIFAGI YHQREKRARDAKTKEELQDGDTSKNSSVLKLNAGGADGNDAYTLSGAVVDTKG GGMGVVFGEYSCYDEKTKQLKEEKLLVELPPVTHAAQTTLMMDGWEACSTSTL DLLKTRPQAHPSGTHIEMVTYNALPLATVLEANTGQASAQGSKSGSFLDNSGQLQ FDYAVQSSDQMSFKAIEEAVKAAASGDLDIMRDDDIPEPPPPPRSFLAAQQQQQQ QQQQQQMGILRQAGAAGGASGSGSSGKKRVHIQEISV
D. pseudoobscura	MDVLHVLDVVDVAQAASLHQPQPQSQSQSETPSKSPTGSGHGQLILTATRTKTRR TTRTRTRSKLRQTKIIQSRFLHIYGYFFFGSLMSYIKTICASMVQQMAAEADKETE VEVAVALPAAKAKRKPVTETETATGMGVGMGTGTDARIVATGICIIRLITLKRFLE ISTTDKHNQKHQQQQHQHQPWQSAASAMSPTSVEASAVEAGGGGEADPVQNRR HSHPHTPSHSHIRHPQACWRHQLTPGILLMLLLLSSLWPDCCDCLHGGGNTVKTK YGLLRGIVVRSTPLVEAFLGIPYASAPVGSLRFMPPITPSTWKTVRNADRFSPVCPQ NVPIPPNGPEALLEVPRARLAQLRRLLPLLKNQSEDCLYLNIYVPYETRRSKRNVD ADVESAETTTTKLSTVVFIHGESFDWNSGNPYDGSELAAHGNVIVVTINFRLGIFG FLKTGGKESAQGNFGLMDLVAGLHWLKENLPAFGGDPQSITLLGYGTGAVLANI LVVSPVSSDLIQRTVLVSGSALSPWAIQKNPLFVKRRVAEQTGCHGDMLYDDLAP CLRTKSVAELLAVKVDHPRFLVGFAPFVDGTVISAVANPLGSTTLPLGSAIVSTSGI EYANFPKRDLIFCLSSVESYLDLSAQDLEFGFNETRRDRILRTFVRNNFHYHLNEIF AVLKNEYTDWEKAIRNPLSSRDATLQFLSDGHTASPLIKLGYMHSLRGGRTYFLH FKHKTVEEEYPQRTGSVRGEDVPFWLGLPISPLFPHNYTTQERQIGRLMLRYLSNF AKTGNPNQSTEKTALGNAEDATLHQQKKRSAAVVPNPHLSDAMTLAVLYNQRR SNAMHEKRSYIRRRLRSNNDGAFTQLGLASERDVGSYEGDDLPFWDAYDVVNQL YVELGNKANIQSHYRGHKLSMWLNLIPQLHRHFNINDQSMRHHQFQDDMNNRD LYEGVVRPQLQTKPAEDDNIIIIQKSRTTTTTPPPPKTPSANVTQALNPMISGTATTT ECGIDGAMSVSELTTTQAPKDNRTGVHTRVETQKDLATASTGIIGNLELLRRLSGK QFQSYTTALIATVAVGCFLLILNVLIFAGIYHQREKRARDAKTKEELQESDTSKSSS MLKLNALIGGGGGGPGGCGSVDHADAYSLSGTVVDSKGGVGVVFGEYSCYDEK TKQLKEEKLIVELPPSSSVSVSGQTSLMMDSWGACSTSTLDLLKTKPGIGMADSH LEMVSYNALPTLMESTSALSSKRGSFVENSLQFNSPQQFDYAVQSSDQMSFKAIEE AVKAAAIQDSEIMRDDDIPEPPPPPRSFLAAQQQQQQQQQQQQQQGILRQTGTAG SATTGPGSSGSKKRVHIQEISV
D. virilis	MPTHTQQQREQEQQQQSAQRKRHLWHTQPASCWQHQSSRNSNSIHSSWLLLLLL LSSLWPDCECLHGAKLSPSTVKTKYGLLRGIVVRSTPLVEAFLGIPYASPPVGSLRF MPPITPSTWKTARNADRFSPVCPQNVPIPPNGPEALLEVPRARLAQLRRLLPLLKN QSEDCLYLNIYVPYETQRPRPAANASSSTSSSSSSSSGSDAMLSTVVFIHGESFEWN SGNPYDGSELAAHGNVIVVTINFRLGIFGFLKTGGKESAQGNFGLMDLVAGLHWL KENLPAFGGNPQSITLLGYGTGAVLANILAVSPVASDLIQRAVLISGTALSPWAIQ KNPLFVKRRVAEQTGCHGDMLYDDLAPCLRTKSVAELLSVKIDHPRFLVGFAPFV DGTVITPNTDAMSSTALPLGSAIVSTSGIEYANFPKRDLIFCLTSVESYLDLSAQDL EFGFNETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRNPLSARDATLQF LSDGHTASPLIKLGYMHSLRGGRTYFVHFKHRTIEEEYPQRTGSVRGEDVPFWLG LPVSPLFPHNYTAQERQIGRLMLRYLANFAKTGNPNQSAPGSAQLPATSQSRAPV

	HLQRKRSAPKYQGALATGEALNLAVIYNQRRSNAMHEKRSYFRRRLRSNDGPAP QLELDSTSTEHEVGSYDGDELPFWDAYDVVNQLYIELGNKANIQSHYRGHKLSM WLNLIPQLHRHFNINDQSMRHHQFQDDLNSRDLYEGVVRPQLQTKPADDDNIIIIQ TGRTTTNPPPPAQKHENTNATNALSATGATATTECGIDGALSVSELTTTQAPKDN RTQHTQVETHKDLATASTGIIGNLELLRRLSGKQFQSYTTALIATVAVGCFLLILN VLIFAGIYHQREKRARDAKTKEELQDGDTSKNSSILKLNAAGSGDAVGSDAYTLS GAVVDTKGAMGMVFGEYSCYDEKTKQLKEEQLLVELPPQPHPAQTTLLMENWE ACSTSTLDLLKTRPAAAQPSSGTHIEMVSYNALPALVETNPLPGSKRGSFLDTSGQ LQFDYAVQSSDQMSFKAIEEAVKAAASGDSDIMRDDDIPEPPPPPRSFLAAQQQQ QQQQQQQMGILRQAGAAGGASSSGGGKKRVHIQEISV
D. willistoni	MATTETEIDETDADADTYGNVDVDMEMEVIADTETDARTMMMGICIIRLITLKRF LEISSSTHKRKTNINDNNDDKEDENDDENEVEENEEMEEEEKENDKHNQWQSAK RRHQLYTHPQSCWRHQLTPSTMIMLLLLLSSHWPDCCQCLQGGSNTVKTKYGLL RGIVVRSTPSPLVEAFLGIPYASPPVGSLRFMPPITPSTWKTVRNADRFSPVCPQNV PIPPNGPEALLELPRSRLAQLRRLLPLLKNQSEDCLYLNIYVPYETQHLRRYTSERP NQETDSTTAEDSAEDSSPKLSTVVFIHGESYDWNSGNPYDGSELAAHGNVIVVTIN FRLGIFGFLKTGGKESAQGNFGLMDLVAGLHWLKENLPAFGGDPQSITLLGYGTG AVLANILVVSPVASDLIQRTVLISGSALSPWAIQKNPLFVKRRVAEQTGCHGDML YDDLAPCLRTKSVAELLAVKIDHPRFLVGFAPFVDGTVISPGANPMGSTTLPMGS AIVSTSGIEYANFPKRDLIFCLTSVESYLDLSAQDLEFGFNETRRDRILRTFVRNNF HYHLNEIFAVLKNEYTDWEKAIRNPLSSRDATLQFLSDGHTASPLIKLGYMHSLR GGRTYFLHFKHKTIEEEYPQRTGSVRGEDVPFWLGLPVSPLFPHNYTTQEHQIGRL MLRYLSNFAKTGNPNQSTANAPLPQQGLSRDASVATSHQLQKRSTVHASGASEA MSLAVLYNQRRSNAMHEKRGYVRRRLRSNDAGAGVFTQLGLGNGGDTGSYDG DELPFWDAYDVVNQLYVELGNKANIQSHYRGHKLSMWLNLIPQLHRHFNINDQS MRHHQFQDDMNNRDLYEGVVRPQLQIKPADDDNILLIQSTRTTTPPPAPSKNHSP NATQAVNAITTGVTTTECNIDGAMSVSELTTTQAPKDNRTGQTTRLENQKDLATA STGIIGNLELLRRLSGKQFQSYTTALIATVAVGCFLLILNVLIFAGIYHQREKRARD AKTKEELQEGDTSKNSSVLKLNGDGSDAYTLSGAVVDSKSGGSGAVAGVGVIFG EYSCYDEKTKQMKEEKLLVELPPASTSSAGQTSLMMDNWGPCSTSTLDLLKTKPI EMVTYNPLPTVIETTSGLSSKRGSFVLDPGMQFPSPQQFDYAVQSSDQMSFKAIEE AVKAGVTSEADIMRDDDIPEPPPPPRSFLAAQQQQQQQQHQQQTGILRQTTSGGG TVPSSGTSGGKKRVHIQEISV
Ceratitis capitata	MLRIRHAVLENPMKAALTEAAAQRFLQIPRQQSNVLFHKRLQKINCQTGKFVYTS NALLPSWLRLALRLCKRRRIYSVEEKPKTWKQAKSKIVTVKAETTRKPCTVDKEA TTTVTTSVEVLELSLQTMPSTPTQPVESTESASAQTQLQLQTCIPQSVSEEKHLLSP PTKQPLSLPPTLLATTETQPQPLSPTKPQALAQSQPQLMCQLEQQPQQQKQEQQL HTAEKQQQQQDEPQVQQQHQQQQQQQTQHSQLQMSKEKHELTCGNTQPHLNSS ATSLSSSPSTSSLSSLSSSSTVLAQFFCHFHLTDYVRDHCRWRWRCWPCHHRPFLP LLHRTLTVLILLSSLWHGTECFHGSTVRLSSNTVKTKYGLLRGIVVRSSPMVEAYL GIPYASPPVGSLRFMPPITPSTWKNVRSADRFSAVCPQTVPIPPNGPEALLEVPRAR LAQLRRLLPLLSNQSEDCLYLNIYVPYENRRKKRNVEHTHDPPTTQLATVLLIHGE SYEWNSGNPYDGSELAAHGNIIVVTINFRLGIFGFLKTGGKESAQGNFGLMDLVA GLHWLKENLPAFGGDPQRITLMGHSTGAALANILVVSPVASDLIQRAVLVSGSAL SPWAIQKNPLFVKRRVAEQTGCHGDMLYDDLAPCLRTKSIAELLAVKIDHPRFLS GFAPFIDGTVISPSSDSVGELSLPLGSAIVSTSGIEFANFPKRDLIFCLTSVESYLDLS AQDLEFGFNETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRSPLSSRDA TLSFLSDGHTAAALIKLGYMHSLRGGRNYFIHFKHRTMEEEYPQRTGSVRGEDVP FWLGLPISPLFPHNYTTQEKQISRLMLRYLANFAKTGNPNHSGLDLPSISPSTSRST GYTTQLLDHGKVKRASLKSLNAVNKFFNLTLPHQLAAQHGNLLDAAGAINLALI YNQRRSGAMRERRAYYKRHLHNNNNNNNNNNNNESNELLNMALRFAADDPLP NGRGEHLPFWDAYDVVNQLYLELGNKAEVQSHYRGHKLSMWLNLIPQLHKHSN INDQSMRHHQFPDDISNRDLYEGVVRPQLQIKPPDDEDTTAAKTPKSTAATKVVS NNTKTGGNVLVGATTTECNPDIPVYTNVGDSASTQQPSVHTDNRSLLENGGQQK

	DLATASTGIIGNLEMFRRLSGKQFPSYTTALIATVAVGCFLLTLNILIFAGIYHQRE KRARDAKTKEELQDTDTSKNSSILKLNAIADDDVSGMHFDGGISGKTTVVFGGEY NYYDEKPKCGKDEKLHVELPPTQTTAHMELNWPGGSSSTLDLLKTKHHSALEAP SYSGSRDVTNELQQLQAISNIEMATYSAQMPQMPIGLAVLDSSSVVVPHGSMRRN SFVASGSQTSMQFDYAVQSSDQLSFKEIENAVKVSNDDIIQDLVLDDDDIPEPPPPP RSFQGVHHHQQQQQHPQQQQQGGILRSSGSNTVSSTSGKKRVHIQEISV
Aedes aegypti	MQKIPTSYHLGQDEVHFTPCSIHPSHQHLIQLSSEPSFIVSGKESFITTIQLMCHFFLC FVVSCATTRVNLVYSNQAMASNFITHHTDERKQSQCFKSILIKTYLFPRKTALRISS IFTFVNTTLEYHSSRTYKFCSTVLNKPSESFTIESNVKTKSDEFNVVMRLSDEISDK MDSMVALQSEVVVQGRKSCIVLKVYSRTTKYKYDCVQYDKSFSTYHSIEHTELII YNQLKHIVNGIDYMRKKFTIHIYNNSTRTYDVFYIVDWIDYRIYIAAKNDNCCNIV HCCSCDRYKEIVFALGVGCTRRTTERTHFYHIKQNAGSRRKNATLSHILETTTTIFT MTTWMLLLMRKWKNIQINLQMLLVLSLCVPIVSAAYQNNKYSTNIIKTKYGPLR GIVMHSNPIVEAFLGVPYASPPIGSLRYMPPVTPSTWKATKLADNFSPVCPQALPK LYGSDGLFEHTRGRLAHLRRLLPLLSNQSEDCLYLNLYVPRSGESVEPDGTTKATI VYIHGESYEWNSGNPYDGSILASTGNVILVTINFRLGVLGFLKTGAKGSAQGNFGL MDLVAALHWLRENLSAFHGDPSRITLMGHGTGAALANILVVSPVASDLIHRVVLL SGSALSPWAIQRDPLAIKQYVANQTGCQLDLLADDIAPCLRTKAVTELLRITPPNP RFLPGYAPFVDGTVIINPRTANIRLPTLPLGSAITSTNGIEFANFPRVELLFGLTSYES YNDLNAQDLEFGFNETRRDRILRTYVRNVFHFHLKEIYSTLRNEYTDWEHPPRNL LGHRDTILELLSDGHTAAPLVRLGYLHSLQEGKSYFLHFRHQSGERDFPERGGSV RGEDVPFTFGLPVSPLFSSNYSLEDKQISQILVQYLTNFAKTGMPNGLKTPNVRDIS KNDHGSIRSLYNVNSIPRYKRSNLKFRLQKDQIPPLLLNNKLLRSLYRSNMNTSYF DTNSNEEYNSNKDNDSWDDTINEPDEINVMQDSPSYGRRESYDENNSDDSSDDN DSFDDIKKDTVLKLENWDMYDSINQIYLELGCSVTARSHYRGHKLSMWLSLIPQL HSFDDAAYLPMRHHHFTDDKPEYYDGRLKEPPNITLPRLITIAKTTTKPQKTETNV VPFAQTVPTECPPNMTVMPPIAATYPQRNQMNTIDSSKDIINQLANSQYQNYSTAF NITIGVGCFLLLLNVIIFSAIYYQREKHANLNKQKENAISEDALHDESTSTEGRFEQ QKIVRSFEKDDTGSKSNIKFQCVSSGASLNDYSCFEQGSDLKKKCIIVDVCSSELPL KEYPYTSPRGSTTGSIHRTVTPEIYKNLSHENLTAVPQAYSSQPSSMSDSSSITVSTS HPMRHNASTTTGDQCTQSEQACLPETQEIGTTVNEVDLDYSSMMMETSQATRSA GFHGGILRQQTGSVIHGTAKKRVQIQEISV
Anopheles gambiae	MQAHDGPKRSYRHDRQQQQQPEPEVSSRVTSCAGFVLQTKLVPARATIWLWVL WCACSFAVAPARAGPRYSSRIVDTKSGAIRGVILELNSKHLEPVEVFKAVPYATPP IGSLRFEPPKKLPPWKGTKLADTFGSVCPQSFPDISNRTAALLSMPKGRYQHLKRL QPLLANQSEDCLTLNIYVPGSGSRGLEAPYSIFFYIHGEAYDWGSGNPYDGSVLAS YGHVIVVTVNFRLGILGFLKTRASLSPGSGGNLGLMDIILALQWVRDNIASFGGDP KRITIVGHDTGAALANLALISKAGKGLVQRAILLSGSALSPWALIPDPDAVRLEVS QQMACHLVPGRNGRKPTTDDITECLRDKPIEALMGVRLTSVRFMPSWGPFLPLED SMDPEFAMEHSGEGFITSELMLGMTTTESYSDFSASDIQYGLEEDQRNRLLRTYIR NAFTFHLNEIFSAVRNEYTDWDKPIQHPINIRDSTMEALSDGHTVAPIIKVAYLHA RRGAKTYMFHFGYQSKESEYPQRLGSVRGEDLPYIFGLPLVQGGPVFPQNYSRQD MGVNEAVLNFVTNFCKTGDPNEAGHQQAAPLHPDYGTAKERTRFRLITWETYET TTQQYLSISTKPKMRSHYRGHKMALWLNLIPQLHRPGDPEVSMRHHHFREREPH YYAGKFQETVSSTMRRNRARNRSAPSAHRTRRWARCCRKMASGAAVAAVPATP SRTCSARRRRRSCWRSSYTAALGVTVGVGCLLLLLNMLIFAGIYYQRDRTKRKSQ TSQSSGTVGGKIPLTSIPSPSPVKAKRTIEPPPSYATLPKRTGNNGASTGIHHHPPTV LGQHQHQHHHHHHHAHHHASGSGSVTTGSGPAGGGPAAPGSTATLKKRVQIQE VTV
Culex quinquefasciatus	MVGCLEDSVEPSGTLKATIVYIHGESYEWNSGNPYDGSVLAAEGNVILVTINFRL GVLGFLKTGAKGSAQGNFGLMDLVAALHWLTENLSAFHGDPSQITLMGHGTGA ALANILVVSPVASDLIHRVILLSGSALSPWAIQREPLAVKRVVAEQTGCRLNVLTE DLAPCLRTKSVSELLEIVQENPSTNGIEFANFPTRQLLFGLTSYESFNDLSAQDLEF

	GFNETRRDRILRTYVRNIFHIHLKEIYSALKNEYTDWEHTPRNPLGYRDTILELLSD GHTAAPERAVVNSGDESDEYNNNWNAFLRPLADKQSIDSEEHNDGDDDDITRDE SEKALLRLENWETYDAINQIYMEIDDSAYLPMRHHHFTEVKPEYYDECPPNMTV MPPVAATHPQRNQMNPSTQKDSAIQSDNSQLDTNSTALKITVCVGCFLLLLNVFIF SAIYYQREKHANLNKQRDTAADTSSHSPLSSLDDQFMQQKSLTSLRAEHSSSTSNI NCSESNTFEKHNCYEKQACTKKKCILVDVSRSDLQMKECSYASPRGSTSGSIQRST TPETYKEMLNENVTATATFSSQPSSLSDSSSITTISCLQMRKNSSTSTCAPETQEIGT TVDEADLEFSALLMEQPSAPTRSIGFQQGGILRQQASSINHGSAKKRVQIQEISV

Figure 7. Neuroligin 2 Homologue Alignment

```
CLUSTAL 2.1 Multiple Sequence Alignments


Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: Dmel        1248 aa
Sequence 2: Dana        1249 aa
Sequence 3: Dgri        1253 aa
Sequence 4: Dpso        1350 aa
Sequence 5: Dvir        1144 aa
Sequence 6: Dwil        1234 aa
Sequence 7: Ccap        1488 aa
Sequence 8: Aaeg        1433 aa
Sequence 9: Agam         885 aa
Sequence 10: Cqui        667 aa
```

---

clustalw.aln

```
CLUSTAL 2.1 multiple sequence alignment


Aaeg            -----------MQKIPTSYHLGQDEVHFTPCSIHPSHQHLIQLSSEPSFIVSGKESFITT
Cqui            ------------------------------------------------------------
Dgri            ------------------------------------------------------------
Dvir            ------------------------------------------------------------
Dmel            ------------------------------------------------------------
Dana            ------------------------------------------------------------
Dpso            ----------------------------------MDVLHVLDVVDVAQAASLHQPQPQSQ
Dwil            ------------------------------------------------------------
Ccap            MLRIRHAVLENPMKAALTEAAAQRFLQIPRQQSNVLFHKRLQKINCQTGKFVYTSNALLP
Agam            ------------------------------------------------------------


Aaeg            IQLMCHFFLCFVVSCATTRVNLVYSNQAMASNFITHHTDERKQSQCFKSILIKTYLFPRK
Cqui            ------------------------------------------------------------
Dgri            -------------------------MTTTNLRSKKQQPTKLKLNQSRFLRIYCYVLVGS
Dvir            ------------------------------------------------------------
Dmel            ---------------------------------------MAQTKIIQSRFLHIYGLLCLGS
Dana            -------------------------------------------------MTTLHMNQKLNMRT
Dpso            SQSETPSKSPTGSGHGQLILTATRTKTRRTTRTRTRSKLRQTKIIQSRFLHIYGYFFFGS
Dwil            ------------------------------------------------------------
Ccap            SWLRLALRLCKRRRIYSVEEKPKTWKQAKSKIVTVKAETTRKPCTVDKEATTTVTTSVEV
Agam            ------------------------------------------------------------


Aaeg            TALRISSIFTFVNTTLEYHSSRTYKFCSTVLNKPSESFTIESNVKTKSDEFNVVMRLSDE
Cqui            ------------------------------------------------------------
Dgri            LVSYIRAICATTVKKLQMTAETEMEIR------------------------SISVSNSNS
Dvir            ------------------------------------------------------------
Dmel            LMGCIKTIGASIAQQTLADADAPDEAKAKEREAETDADADAEGDAKAETETGKGTREHTN
Dana            TAVKMKRTRYPASCPSRSQTVAEADGADKEREAEPETGAGARAEAGSREGAEAKRRAETE
Dpso            LMSYIKTICASMVQQMAAEADKETEVEVAVALPAAKAKR------KPVTETETATGMGVG
Dwil            ------------------------MATTETEIDETDADA------DTYGNVDVDMEMEVI
Ccap            LELSLQTMPSTPTQPVESTESASAQTQLQLQTCIPQSVSEEKHLLSPPTKQPLSLPPTLL
Agam            ------------------------------------------------------------


Aaeg            ISDKMDSMVALQSEVVVQGRKSCIVLKVYSRTTKYKYDCVQYDKSFSTYHSIEHTELIIY
Cqui            ------------------------------------------------------------
Dgri            DSNSDSSLQLRTRICIIRLITLKRFLHILAMPTHNQQQQQQQ------------------
Dvir            ------------------------------------------------------------
Dmel            GDAEADARIMATGICIIRLITLKRFLENSKTEEHNQRRQQ--------------------
Dana            ADAEADARIMATGICIIRLITLKRFLEISTTNEHNQRRQQ--------------------
Dpso            MGTGTDARIVATGICIIRLITLKRFLEISTTDKHNQKHQQQQHQHQPWQSAASAMSPTSV
Dwil            ADTETDARTMMMGICIIRLITLKRFLEISSSTHKRKTNINDNNDDKEDENDDENEVEENE
Ccap            ATTETQPQPLSPTKPQALAQSQPQLMCQLEQQPQQQKQEQQLHTAEKQQQQQDEPQVQQQ
```

```
Agam            ------------------------------------------------------------


Aaeg            NQLKHIVNGIDYMRKKFTIHIYNNSTRTYDVFYIVDWIDYRIYIAAKNDNCCNIVHCCSC
Cqui            ------------------------------------------------------------
Dgri            QQHHQHQHHQQQQQLQSAHRKRHLWHSQVLTQAPSCWRHQTS------------------
Dvir            MPTHTQQQREQEQQQQSAQRKRHLWHTQPASCWQHQSSRNSNS------------------
Dmel            SATSAASAAPAGATSPRRRHLHYPGHSAVQP----------GH------------------
Dana            SATSAGVTSPRSRRRHPHSHPGQPGHTASHHTAGHPEACWRHQ------------------
Dpso            EASAVEAGGGGEADPVQNRRHSHPHTPSHSHIRHPQACWRHQL------------------
Dwil            EMEEEEKENDKHNQWQSAKRR-------HQLYTHPQSCWRHQL------------------
Ccap            HQQQQQQQTQHSQLQMSKEKHELTCGNTQPHLNSSATSLSSSPSTSSLSSLSSSSTVLAQ
Agam            MQAHDGPKRSYRHDRQQQQQPEPEVSSRVTSCAGFVLQTKLVP------------------


Aaeg            DRYKEIVFALGVGCTRRTTERTHFYHIKQNAGSRRKNATLSHILETTTTIFTMTTWMLLL
Cqui            ------------------------------------------------------------
Dgri            ------------------------------------------------------------
Dvir            ------------------------------------------------------------
Dmel            ------------------------------------------------------------
Dana            ------------------------------------------------------------
Dpso            ------------------------------------------------------------
Dwil            ------------------------------------------------------------
Ccap            FFCHFHLTDYVRDHCRWRWRCWPCHHRPF--------------------------------
Agam            ------------------------------------------------------------


Aaeg            MRKWKNIQINLQMLLVLSLCVPIVSAAYQNNKYSTNIIKTKYGPLRGIVMHSNP----IV
Cqui            ------------------------------------------------------------
Dgri            -----GSTWLLLLLLLSSLWP--DCECVHGAKLSANTVKTKYGLLRGIVVRSTP----LV
Dvir            ----IHSSWLLLLLLLSSLWP--DCECLHGAKLSPSTVKTKYGLLRGIVVRSTP----LV
Dmel            ----PEACALLVLLLLTSLWP--DCCECLHGGS--NTVKTKYGLLRGIVVRSSP----LV
Dana            ----LTPSTLLMLLLLTSLWP--DCCDCLHGSS--NTVKTKYGLLRGIVVRSSP----LV
Dpso            ----TPG-ILLMLLLLSSLWP--DCCDCLHGGG--NTVKTKYGLLRGIVVRSTP----LV
Dwil            ----TPSTMIMLLLLLSSHWP--DCCQCLQGGS--NTVKTKYGLLRGIVVRSTPS--PLV
Ccap            --LPLLHRTLTVLILLSSLWHGTECFHGSTVRLSSNTVKTKYGLLRGIVVRSSP----MV
Agam            ------ARATIWLWVLWCACSFAVAPARAGPRYSSRIVDTKSGAIRGVILELNSKHLEPV


Aaeg            EAFLGVPYASPPIGSLRYMPPVTPSTWKATKLADNFSPVCPQALPK-LYGSDGLFEHTRG
Cqui            ------------------------------------------------------------
Dgri            EAFLGIPYASPPVGSLRFMPPITPSTWKTTRNADRFSPVCPQSVPIPPNGPEALLEVPRA
Dvir            EAFLGIPYASPPVGSLRFMPPITPSTWKTARNADRFSPVCPQNVPIPPNGPEALLEVPRA
Dmel            EAFLGIPYASPPVGSLRFMPPITPSTWKTVRSADRFSPVCPQNIPIPPNGPEALLEVPRA
Dana            EAFLGIPYASPPVGSLRFMPPITPSTWKTVRSADRFSPVCPQNIPIPPNGPEALLEVPRA
Dpso            EAFLGIPYASAPVGSLRFMPPITPSTWKTVRNADRFSPVCPQNVPIPPNGPEALLEVPRA
Dwil            EAFLGIPYASPPVGSLRFMPPITPSTWKTVRNADRFSPVCPQNVPIPPNGPEALLELPRS
Ccap            EAYLGIPYASPPVGSLRFMPPITPSTWKNVRSADRFSAVCPQTVPIPPNGPEALLEVPRA
Agam            EVFKAVPYATPPIGSLRFEPPKKLPPWKGTKLADTFGSVCPQSFPDISNRTAALLSMPKG


Aaeg            RLAHLRRLLPLLSNQSEDCLYLNLYVPR----------------------SGESVEPDGTT
Cqui            ------------------------MVGC----------------------LEDSVEPSGTL
Dgri            RLAQLRRLLPLLKNQSEDCLYLNIYVPEEARRTQPDN-------EDADSSSSSSSSSDSL
Dvir            RLAQLRRLLPLLKNQSEDCLYLNIYVPYETQRPRPAANA----SSSTSSSSSSSSGSDAM
Dmel            RLAQLRRLLPLLKNQSEDCLYLNIYVPYETRRQRR-------------NTDDTTGEPKTK
Dana            RLAQLRRLLPLLKNQSEDCLYLNIYVPYETRRQRR-------------NIDDTNGEPKTK
Dpso            RLAQLRRLLPLLKNQSEDCLYLNIYVPYETRRSK----------RNVDADVESAETTTTK
Dwil            RLAQLRRLLPLLKNQSEDCLYLNIYVPYETQHLRRYTSERPNQETDSTTAEDSAEDSSPK
Ccap            RLAQLRRLLPLLSNQSEDCLYLNIYVPYENRRKKR-------------NVEHTHDPPTTQ
Agam            RYQHLKRLQPLLANQSEDCLTLNIYVPG----------------------SGSRGLEAP
                                          *                                .

Aaeg            KATIVYIHGESYEWNSGNPYDGSILASTGNVILVTINFRLGVLGFLKTGAKGS--AQGNF
Cqui            KATIVYIHGESYEWNSGNPYDGSVLAAEGNVILVTINFRLGVLGFLKTGAKGS--AQGNF
Dgri            LSTVVFIHGESYEWNSGNPYDGSELAAHGNVIVVTINFRLGIFGFLKTGGKES--AQGNF
Dvir            LSTVVFIHGESFEWNSGNPYDGSELAAHGNVIVVTINFRLGIFGFLKTGGKES--AQGNF
Dmel            LSTVVFIHGESYDWNSGNPYDGSELAAHGNVIVVTINFRLGIFGFLKTGGKES--AQGNF
Dana            LSTVVFIHGESYDWNSGNPYDGSELAAHGNVIVVTINFRLGIFGFLKTGGKES--AQGNF
Dpso            LSTVVFIHGESFDWNSGNPYDGSELAAHGNVIVVTINFRLGIFGFLKTGGKES--AQGNF
Dwil            LSTVVFIHGESYDWNSGNPYDGSELAAHGNVIVVTINFRLGIFGFLKTGGKES--AQGNF
```

```
Ccap            LATVLLIHGESYEWNSGNPYDGSELAAHGNIIVVTINFRLGIFGFLKTGGKES--AQGNF
Agam            YSIFFYIHGEAYDWGSGNPYDGSVLASYGHVIVVTVNFRLGILGFLKTRASLSPGSGGNL
                 : .. ****:::*.******** **: *::*:**:*****::***** .. *  : **:

Aaeg            GLMDLVAALHWLRENLSAFHGDPSRITLMGHGTGAALANILVVSPVASDLIHRVVLLSGS
Cqui            GLMDLVAALHWLTENLSAFHGDPSQITLMGHGTGAALANILVVSPVASDLIHRVILLSGS
Dgri            GLMDLVAGLHWLKENLPAFGGNPQSITLLGYGTGAVLANILAVSPVASDLIQRAVLISGT
Dvir            GLMDLVAGLHWLKENLPAFGGNPQSITLLGYGTGAVLANILAVSPVASDLIQRAVLISGT
Dmel            GLMDLVAGLHWLKENLPAFGGDPQSITLLGYGTGAVLANILVVSPVASDLIQRTVLVSGS
Dana            GLMDLVAGLHWLKENLPAFGGDPQSITLLGYGTGAVLANILVVSPVASDLIQRTVLVSGS
Dpso            GLMDLVAGLHWLKENLPAFGGDPQSITLLGYGTGAVLANILVVSPVSSDLIQRTVLVSGS
Dwil            GLMDLVAGLHWLKENLPAFGGDPQSITLLGYGTGAVLANILVVSPVASDLIQRTVLISGS
Ccap            GLMDLVAGLHWLKENLPAFGGDPQRITLMGHSTGAALANILVVSPVASDLIQRAVLVSGS
Agam            GLMDIILALQWVRDNIASFGGDPKRITIVGHDTGAALANLALISKAGKGLVQRAILLSGS
                ****:: .*:*: :*:.:* *:*. **::*:.***.***:  :* ....*::*.:*:**:

Aaeg            ALSPWAIQRDPLAIKQYVANQTGCQLD-------LLADDIAPCLRTKAVTELLRITPPNP
Cqui            ALSPWAIQREPLAVKRVVAEQTGCRLN-------VLTEDLAPCLRTKSVSELLEIVQENP
Dgri            ALSSWGIQKNPLFVKRRVAEQTGCHGD-------MLYDDLAPCLRTKSVAELLAVQIDHP
Dvir            ALSPWAIQKNPLFVKRRVAEQTGCHGD-------MLYDDLAPCLRTKSVAELLSVKIDHP
Dmel            ALSPWAIQKNPLFVKRRVAEQTGCHGD-------MLYDDLAPCLRTKSVAELLAVKVDHP
Dana            ALSPWAIQKNPLFVKRRVAEQTGCHGD-------MLYDDLAPCLRTKSVAELLAVKVDHP
Dpso            ALSPWAIQKNPLFVKRRVAEQTGCHGD-------MLYDDLAPCLRTKSVAELLAVKIDHP
Dwil            ALSPWAIQKNPLFVKRRVAEQTGCHGD-------MLYDDLAPCLRTKSVAELLAVKIDHP
Ccap            ALSPWAIQKNPLFVKRRVAEQTGCHGD-------MLYDDLAPCLRTKSIAELLAVKIDHP
Agam            ALSPWALIPDPDAVRLEVSQQMACHLVPGRNGRKPTTDDITECLRDKPIEALMGVRLTSV
                ***.*.:  :*  ::  *::* .*:            :*:: *** *.:  *: :

Aaeg            RFLPGYAPFVDGTVIINPRTANIRLPTLPLGSAITSTNGIEFANFPRVELLFGLTSYESY
Cqui            -----------------------------------STNGIEFANFPTRQLLFGLTSYESF
Dgri            RFLVGFAPFVDGTVIS-PNTETMSSTALPLGSAIVSTSGIEYANFPKRDLIFCLTSVESY
Dvir            RFLVGFAPFVDGTVIT-PNTDAMSSTALPLGSAIVSTSGIEYANFPKRDLIFCLTSVESY
Dmel            RFLVGFAPFVDGTVIS-PGANPLGSTTLPLGSAIVSTSGIEYANFPKRDLIFCLTSVESY
Dana            RFLVGFAPFVDGTVIS-PGANPLGSTTLPLGSAIVSTSGIEYANFPKRDLIFCLTSVESY
Dpso            RFLVGFAPFVDGTVIS-AVANPLGSTTLPLGSAIVSTSGIEYANFPKRDLIFCLSSVESY
Dwil            RFLVGFAPFVDGTVIS-PGANPMGSTTLPMGSAIVSTSGIEYANFPKRDLIFCLTSVESY
Ccap            RFLSGFAPFIDGTVIS-PSSDSVGELSLPLGSAIVSTSGIEFANFPKRDLIFCLTSVESY
Agam            RFMPSWGPFLP-----------------LEDSMDPEFAMEHSGEGFITSELMLGMTTTESY
                                             :   .  .*   :*:: ::: **:

Aaeg            NDLNAQDLEFGFNETRRDRILRTYVRNVFHFHLKEIYSTLRNEYTDWEHPPRNLLGHRDT
Cqui            NDLSAQDLEFGFNETRRDRILRTYVRNIFHIHLKEIYSALKNEYTDWEHTPRNPLGYRDT
Dgri            LDLSAQDLEFGFNETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRNPLSARDA
Dvir            LDLSAQDLEFGFNETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRNPLSARDA
Dmel            LDLSAQDLEFGFNETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRNPLSSRDA
Dana            LDLSAQDLEFGFNETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRNPLSSRDA
Dpso            LDLSAQDLEFGFNETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRNPLSSRDA
Dwil            LDLSAQDLEFGFNETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRNPLSSRDA
Ccap            LDLSAQDLEFGFNETRRDRILRTFVRNNFHYHLNEIFAVLKNEYTDWEKAIRSPLSSRDA
Agam            SDFSASDIQYGLEEDQRNRLLRTYIRNAFTFHLNEIFSAVRNEYTDWDKPIQHPINIRDS
                 *:.*.*:::*::* :*:*:***::** *  **:**::.::******::. :  :. **:

Aaeg            ILELLSDGHTAAPLVRLGYLHSLQEGKSYFLHFRHQSGERDFPERGGSVRGEDVPFTFGL
Cqui            ILELLSDGHTAAPER-----------------------------------------------
Dgri            TLQFLSDGHTAAPLIKLGYMHSLRGGRTYFVHFKHRTIEEEYPQRTGSVRGEDVPFWLGL
Dvir            TLQFLSDGHTASPLIKLGYMHSLRGGRTYFVHFKHRTIEEEYPQRTGSVRGEDVPFWLGL
Dmel            TLQFLSDGHTASPLIKLGYMHSLRGGRAYFLHFKHKTIEEEYPQRSGSVRGEDVPFWLGL
Dana            TLQFLSDGHTASPLIKLGYMHSLRGGRTYFLHFKHKTIEEEYPQRTGSVRGEDVPFWLGL
Dpso            TLQFLSDGHTASPLIKLGYMHSLRGGRTYFLHFKHKTVEEEYPQRTGSVRGEDVPFWLGL
Dwil            TLQFLSDGHTASPLIKLGYMHSLRGGRTYFLHFKHKTIEEEYPQRTGSVRGEDVPFWLGL
Ccap            TLSFLSDGHTAAALIKLGYMHSLRGGRNYFIHFKHRTMEEEYPQRTGSVRGEDVPFWLGL
Agam            TMEALSDGHTVAPIIKVAYLHARRGAKTYMFHFGYQSKESEYPQRLGSVRGEDLPYIFGL
                 :. ******.:.

Aaeg            PVS---PLFSSNYSLEDKQISQILVQYLTNFAKTGMPNGLKTPNVRDISKNDHGSIRSLY
Cqui            ------------------------------------------------------------
Dgri            PVS---PLFPHNYTTQERQIGRLMLRYLANFAKTGNPNQSTSSTALAAAATSQTQVATHT
Dvir            PVS---PLFPHNYTAQERQIGRLMLRYLANFAKTGNPNQ-SAPGSAQLPATSQSRAPVHL
Dmel            PMS---PLFPHNYTTQERQIGRLMLRYLSNFAKTGNPNQSTAKSVLPNP--NEVLETALH
Dana            PVS---PLFPHNYTTQERQIGRLMLRYLSNFAKTGNPNQSTARPTLTIP--SEESDVSLH
Dpso            PIS---PLFPHNYTTQERQIGRLMLRYLSNFAKTGNPNQSTEKTALG----NAEDATLHQ
```

```
Dwil            PVS---PLFPHNYTTQEHQIGRLMLRYLSNFAKTGNPNQSTANAPLPQQGLSRDASVATS
Ccap            PIS---PLFPHNYTTQEKQISRLMLRYLANFAKTGNPNHSGLDLPSISPSTSRSTGYTTQ
Agam            PLVQGGPVFPQNYSRQDMGVNEAVLNFVTNFCKTGDPNE----------------------


Aaeg            NVNSIPRYKRSNLKF--------------RLQKDQIPPLLLNNKLLRSLYRSNMNTSYFD
Cqui            ------------------------------------AVVNSGDESDEYNNNWNAFLRP
Dgri            QRKR-------------------------SAPRYLDAMATGEALNLAVIYNQRRSNAMHE
Dvir            QRKR-------------------------SAPKYQGALATGEALNLAVIYNQRRSNAMHE
Dmel            QQKK-------------------------RSTSLTHPNLS-EALNLAVLYNQRRSNAMHE
Dana            QQKK-------------------------RSTGLIHPNLS-EALNLAVIYNQRRSNAMHE
Dpso            QKKR-------------------------SAAVVPNPHLS-DAMTLAVLYNQRRSNAMHE
Dwil            HQLQ-------------------------KRSTVHASGAS-EAMSLAVLYNQRRSNAMHE
Ccap            LLDHGKVKRASLKSLNAVNKFFNLTLPHQLAAQHGNLLDAAGAINLALIYNQRRSGAMRE
Agam            ------------------------------------------------AGHQQAAPLHP
                                                                   .

Aaeg            TNSNEEYNSNKDNDSWDDTINEPDEINVMQDSPSYGRRESYDENNSDDSSDDNDSFDDIK
Cqui            LADKQSIDSEEHND----------------------------------GDDDDITRDES
Dgri            KRSYFRRRLRSNDGVS-----------------------------TQQGLG---TELEPG
Dvir            KRSYFRRRLRSNDGPA-----------------------------PQLELDSTSTEHEVG
Dmel            KRSYIRRRLRSNDAA------------------------------FTQLGISSERDVG
Dana            RRSYIRRRLRSNDAS------------------------------FTQLGISSEKDVG
Dpso            KRSYIRRRLRSNND-------------------------------GAFTQLGLASERDVG
Dwil            KRGYVRRRLRSNDAGA-----------------------------GVFTQLGLGNGGDTG
Ccap            RRAYYKRHLHNNNNNNNNNNNN-------------------ESNELLNMALRFAADDPLP
Agam            DYGTAKERTR-------------------------------------------------
                        .

Aaeg            KDTVLKLENWDMYDSINQIYLELGCSVTARSHYRGHKLSMWLSLIPQLHSFDDAAYLPMR
Cqui            EKALLRLENWETYDAINQIYMEI----------------------------DDSAYLPMR
Dgri            GYDGDALPFWDAYDVVNQLYIELGNKVNIQSHYRGHKLSMWLNLIPQLHRHFNINDQSMR
Dvir            SYDGDELPFWDAYDVVNQLYIELGNKANIQSHYRGHKLSMWLNLIPQLHRHFNINDQSMR
Dmel            SYDGDELPFWDAYDVVNQLYVELGNKANIQSHYRGHKLSMWLNLIPQLHRHFNINDQSMR
Dana            SYDGDELPFWDAYDVVNQLYVELGNKANIQSHYRGHKLSMWLNLIPQLHRHFNINDQSMR
Dpso            SYEGDDLPFWDAYDVVNQLYVELGNKANIQSHYRGHKLSMWLNLIPQLHRHFNINDQSMR
Dwil            SYDGDELPFWDAYDVVNQLYVELGNKANIQSHYRGHKLSMWLNLIPQLHRHFNINDQSMR
Ccap            NGRGEHLPFWDAYDVVNQLYLELGNKAEVQSHYRGHKLSMWLNLIPQLHKHSNINDQSMR
Agam            ----FRLITWETYETTTQQYLSISTKPKMRSHYRGHKMALWLNLIPQLHRPGDPE-VSMR
                      *  *: *:  .* *:.:                              :    .**

Aaeg            HHHFTDDK-------------------PEYYDGRLKEPPNITLPRLITIAKTTTKPQKTET
Cqui            HHHFTEVK-------------------PEYYD------------------------------
Dgri            HHQFQDDLNNRDLYEGVVRPQLQTKPAEDDNILVIQTTRTTTT-PLPPQSPKEENTNATN
Dvir            HHQFQDDLNSRDLYEGVVRPQLQTKPADDDNIIIIQTGRTTTN-PPPP-AQKHENTNATN
Dmel            HHQFQDDMNNRDLYEGVVRPQLQTKPAEDDNIIIMQRS-RTTTPPPPSKNPSTNATQALN
Dana            HHQFQDDMNNRDLYEGVVRPQLQTKPAEDEHIIITQRSRTSTTAPPPPKDATPNATQVLN
Dpso            HHQFQDDMNNRDLYEGVVRPQLQTKPAEDDNIIIIQKSRTTTTTPPPPKTPSANVTQALN
Dwil            HHQFQDDMNNRDLYEGVVRPQLQIKPADDDNILLIQSTRTTTPPPAPSKNHSPNATQAVN
Ccap            HHQFPDDISNRDLYEGVVRPQLQIKPPDDEDTTAAKTPKSTAATKVVSNNTKTGGNVLVG
Agam            HHHFRERE-------------------------------------------------P
                **:* :

Aaeg            NVVPFAQTVPTECPPNMTVMPPIAATYP-----QRNQMNTIDSSKDIINQLA--------
Cqui            -----------ECPPNMTVMPPVAATHP-----QRNQMNPS-TQKDSAIQSD--------
Dgri            ALSATGATATTECGIDGAMSVSELTTTQSPKDN--RTHSKVETRKDLATASTGIIGNLEL
Dvir            ALSATGATATTECGIDGALSVSELTTTQAPKDNR-TQHTQVETHKDLATASTGIIGNLEL
Dmel            PLGT-GTPTTTECGIDGAMSVSELTTTQSPKDNRTGIVR-VETQKDLATASTGIIGNLEL
Dana            PSATGATTTTTDCNIDGAMSVSELTTTQSPKDNQTGRVR-VETQKDLATASTGIIGNLEL
Dpso            -PMISGTATTTECGIDGAMSVSELTTTQAPKDNRTGVHTRVETQKDLATASTGIIGNLEL
Dwil            -AITTG-VTTTECNIDGAMSVSELTTTQAPKDNRTGQTTRLENQKDLATASTGIIGNLEL
Ccap            ---ATTTECNPDIPVYTNVGDSASTQQPSVHTDNRSLLENGGQQKDLATASTGIIGNLEM
Agam            HYYAGKFQETVSSTMRRNRARNRSAPSAHRTRRWARCCRKMASGAAVAAVPATPSRTCSA
                          .           :

Aaeg            -----NSQYQNYSTAFNITIGVGCFLLLLNVIIFSAIYYQREKHANLNKQKENAISEDAL
Cqui            -----NSQLDTNSTALKITVCVGCFLLLLNVFIFSAIYYQREKHANLNKQRD-TAADTSS
Dgri            LRRLSGKQFQSYTTALIATVAVGCFLLILNVLIFAGIYHQREKRARDAKTKEELQDGDTS
Dvir            LRRLSGKQFQSYTTALIATVAVGCFLLILNVLIFAGIYHQREKRARDAKTKEELQDGDTS
Dmel            LRRLGGKQFQSYTTALIATVAVGCFLLILNVLIFAGIYHQREKRARDAKTKEELQEGDNS
Dana            LRRLGGKQFQSYTTALIATVAVGCFLLILNVLIFAGIYHQREKRARDAKTKEELQESDTS
```

Figure 7. Neuroligin 2 Homologue Alignment. The yellow, highlighted regions indicate the transmembrane domain for each protein sequence.

Figure 8. Neuroligin 2 Phylogenetic Tree

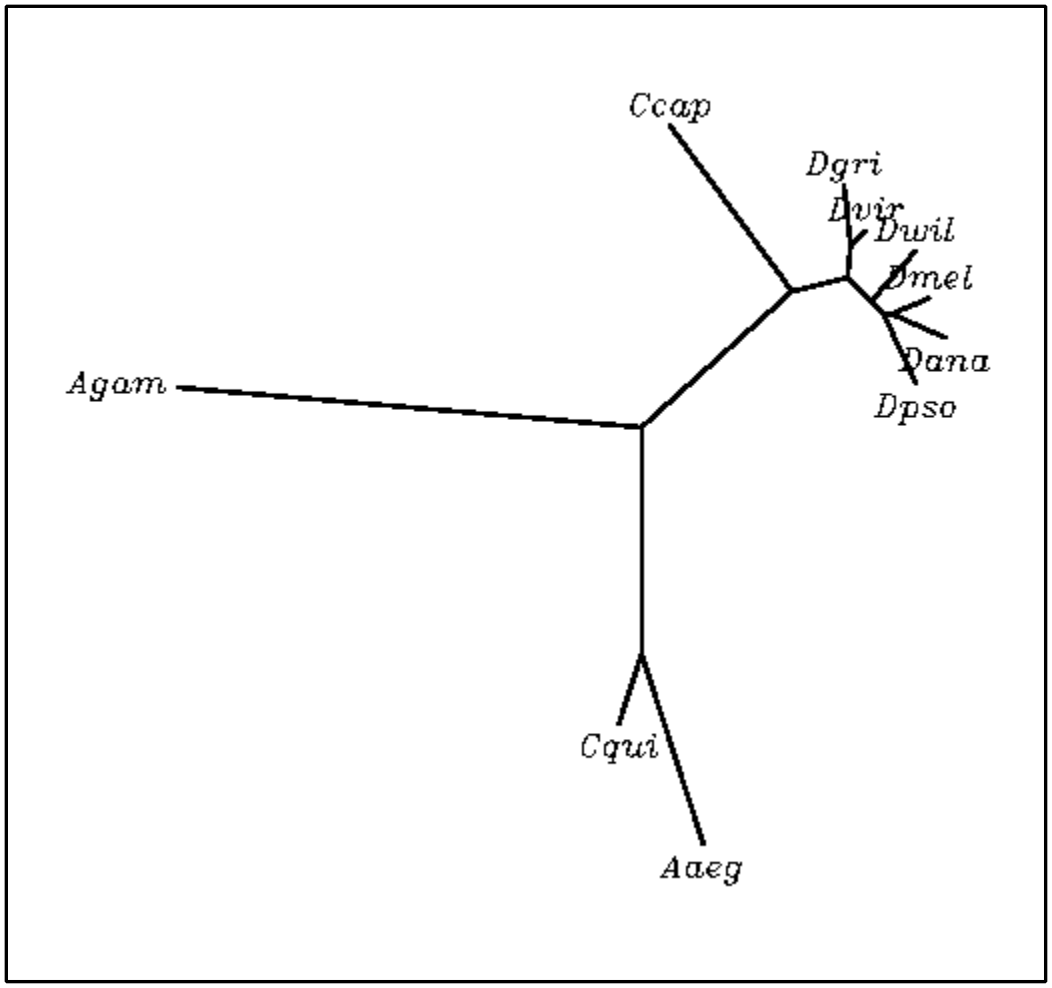


Table 4. Glutactin Homolgues

D. melanogaster	MKPLLLVLALCGAQVHAHSVGLRPDYNDYSDEDTRRDWLPEPLKPVPWQSETRY AQPQEAVVQAPEVGQILGISGHKTIANRPVNAFLGIRYGTVGGGLARFQAAQPIGY QGRVNATVQSPNCAQFPELDRLRLSESRGENVDDCLTLDIYAPEGANQLPVLVFVH GEMLFDGGSEEAQPDYVLEKDVLLVSINYRLAPFGFLSALTDELPGNVALSDLQLA LEWLQRNVVHFGGNAGQVTLVGQAGGATLAHALSLSGRAGNLFQQLILQSGTAL NPYLIDNQPLDTLSTFARLARCPPPSINPSAQGLKPLYDCLARLPTSQLVAAFEQLLL QNEHLGLTQLGGFKLVVGDPLGFLPSHPASLATNSSLALPMIIGATKDASAFIVSRI YDQLARLQSRNVSDYIDVVLRHTAPPSEHRLWKQWALREIFTPIQEQTASLQTVAP GLLELSNYILYRAPVINSISQSYRSVPAYLYTFDYRGEHHRFGHLSNPLPFGVDASLS DDSVYLFPYPPEASRLNPLDRSLSRALVTMWVNFATTGVPNPSSGVWPQATSEYG PFLRFTNNQQSPLELDPHFGEGIYLPNYRVIYKPTTNFSPPITTTTTTTTTTTTTSRPY AYNPYANWQNRPSQQHPNWHPADPEYVRAQEARQQEFIREREQRRREQQLRDQQ RYPQQEPREQQDERIRQQREQEERLRQQREQEERLRQQRELEERIRQQQEREQYER EQQEREQRREEELERQQREREQQQPEQQPEYNPEPVNPWGYPVQEPQPDDNPEDG RLPYPSYEQYGPEGNENLPETDANRNFSEEDREQQQQEQLRREQQEQQEREYQLQ LEREQQEREQQERGQQEPGPEEYPSYEEYSRALQEKNAERDRIYAEEQERERQQQE TLLQENQQHPEQSLPEEQPTHPNYEAYDGDRSYAEEQEREQQRRDQVEQEREEQP DEDQGEEYERSPDEEEAAEQDVLKVEDFPSYEAYLEAATKLREEQEEQEKLEEERY RAQQEEEDRIQAERERNSRN
D. ananassae	MKPLYFVLALCVAQSCALYRGKRGSDYDDDYSDEEESWQPDRVKKEPWKVRHTP PVRVQQDISSQQAPQYAQDSRDVVVQAPEVGEIVGVRGYKSNANGLINAFLGIRY GKVGSGLSRFQAAQPTGYQGRVNANVKSPNCAQFPELQRLRESEARGENVDDCLT LDIYAPEHASNLPVLVFVHGEMLFDGGSEEAQPDYVLEKDVVLVSINYRLAPFGFL SALSDELPGNVALSDLHLALEWLQKNLESFGGNPSQVTLLGQAGGATLAHALSLS GRATGLFQQLVLQSGTALNPYLIDERPLDTLDTFASFAGCPTAANTRSLAPLYDCL GSLPTSNLVAAFEQLFNRNEAIGLSSLGGFKLVVGDPLGYLPKPPVALAANASNAL PMIVGATKDASAFILSRFYNQIEHVRSYNVSDYIDVVLRHLAPPSEHQIWKQWALR EIFSPDQVRTVNSISVSQGLLELSNMILYRAPVIYSIRNSYRNAPVYLYNFAYRGEH HRFEHVKSPLPFAVDASLSDDNVYVFPYPVETSQLNAVDRSLSRALVTMWVNFAT TGVPNPNFGVWPRASSEYGPVLRFSNSRENMLELDPHFGEGINIPNLYSNYFNTTSK TTPAVIPVFTTTTTTTTTTRPYVHNPYNNWNRQIQQTGPSRDAAEEERRKQEFREY QRRQEEERRREYQLRLQQRQEQLQRESLEREKKLQEQREQQQREQELQEQQREQQ QREQQRREQQWREQQQREQQWREQQEREQQERQYGDVNRDPYTVTYDAYGREI TPNSAEQERQKQEREQLEREEQQREQQRREYEERLEQDRQQREREQQEREQQERV QPEGQYQPEENTEDGRQAYPSYEDYLKSQQEEQSPESDPESNYAEEQQREQERLEQ ERQLREREHREYEERLEQERQQREREQQEREEQGGQHQPEENPEDGRDAYPTYED YVRAQQGEQSPESDPDRNYAQEQQREQERLEQERQLREQEHRQYEERLQREREQQ EREQEQPQGQPDNDLSRYPSYDEYLRSQQEGRPSDSDSDPNYDEEQAREQQRRDQ LRQEQEQEYGPSGGYPEESDDQDQEYQPNVDPKNDPSNYATYEDYVKAQLKLRE EEEERDRALREEEERQFAQAEAEQEAARQSQPTSIVFRRFLNAPRHIKVIHRQQGRR H
D. grimshawi	MQLLSLLLLLLVLGICHAQFDEIYSYTPRPLRPAPWLGEVFKTTTPPPPAPPSGGVH VPGVGEVNGLPKYKVIKGRPIDAYLGIRYAQVNSGLGRFQKARAVPYAGRIDATQ MSPNCAQFPEKQRLQEAEARGENADDCLTLNIYAPAGPRELPVLVFVHGEMLFDG SAEEAQPDYVLEHDVVLVSINYRLAPFGFLSALSDELPGNVALSDLQLALEWVQR NIRYFGGASSRVTLIGQAGGATLAHALSLSPQAQHLFQQLILQSGTALNPYLIDERP LETLATFASLARCPTAGRNLAPLYQCLNRMRTSQLVEIFEQLFRNNEPLGLSALGGF KLVVGDRLGYLPNHPAALVASNNSTKPLIVGAAKDASAFILSGFYDQIQNLNSRNL SDYINVILRHTAPPQHHKVWHDWALQEIFTPEQKRYVNTNSVAQGLLELSNLIMY RSPVIDTIRLSHQKTPAYLYTFDYRGQYHRFADVKNPLPFGVDASLSDDSVYLFPYP EETSKLNPDDMSIARALVAMWVSFAQTGIPNPNPNVWPKATSEYGPFLRFTNSKQ

	NILELDQHFGEGINIPNLYPMYFTTSRTSSGTTTTTRRPYSGLPPAYPEYNPRTQEAH WKQQMREQEEQLARENEKRRLEQQQREEQLQQENAQREQLERDQRQRWAAEEL ALKEYELRQQRREREQQQREQHEREQPDATQEQLDPEQEQHSEEQPDQESRWQQER EQQLREQQRREQQQREQRQREQQQREQQQREQQHRQQPEREQEQPAQQQPHYSQ SDVPPEPNDFPTYEDYVKAYEQWVRAAQEQEQHQQSQEERE REAREQEELEQQQ REQQDRQQQERPHHEEEQPEATDEQPEPEQEQHSQTVLDQELRWKQVLEQDEREL RQREEREYEQREQQRRDQQQREQQGREHQEPAQQEPHSSQSDAPPEPNNFPTYED YVKAYEQWAGAAQEQEQQQQSQEDNEQEARDQQEREQQLRQIEQEELAKHASEP EGKPYRYDYSSDTSYMLAVAEWEAAGPDYKPYQRTQEEIEEDARYERKRKILMN MRKDYEEGVQHPREPGAEPRTHEFNNFESFIAAHSKWEEEHPVQKEELESPEQRAQ RDRETRGRVEQERPQHQVLDLELTQPQNGYDNFNDYVNGRALLTAARQDRDQQQ QSQEEREQQAREQYIREQQQRHQLEREQEEQERAQYEPHDDYFLNASPWPEAGQE QDQQQQSQEEYDQEAREQQEREQRERE QLEREEQQRSDAQPDTDTYGLDDSDLN ADLWPQASEEEEADRYPEYSNGHNYPQRVKLMRRHPRLRS
D. pseudoobscura	MQPLPSWVQLLALALCVCQICALQTRPRTDDIDYELSEENEGSFTPQPLRPPPWQE APQSQEAPQEARVQVPGVGHIVGVRGYKTISNRPINAFLGVHYAQVGGGLGRFQQ ARPVPFQGRVDATTASPNCAQFPELQRLQTAEARGENVDDCLTLNIYAPEGARDLP VLVFVHGEMLFDGGAEEAQPDYVLENEVVLVSINYRLAPFGFLSALSEQLPGNVA LSDIHLALEWLQRNLPYFGGNSGKVTLVGQAGGASLVHALSLSGRATNLFQQLILQ SGTALNPYLIDERPLETLDTFARLARCPNAGRSLAPLYDCLSRLPTSQVVSAFEQLF QQNEARGLNFLGGFKLVVGDPLGYLPEHPAALAANSANSTVPMIVGAAKDASAFI LSRFYDQIQRVQSQNVSDFINVVLRHTAPPSQHQLWLDWARREIFSQEHDRTASAY SVSQGLLELSNMILYRAPVVYSIRLSHKKSPVYLYTFDYRGEYHRFGHLDNPLPFG VDASLSDDSVYLFPYPPEASRLNPVDKSLSRALVTMWVNFAVTGIPNQNPGVWPK ATSEYGPFLRFTNSRQNMLELDPHFGDGLYAPNLYGQYFNTTNSTTTATTTTTTTT TRRPYVYNPYPSWQQSPYGYNTTTTTSTTTTTTRRPTYAYNPYANWQQNRQPQPP QAFPRRPTDPAYERAQEIRRQQVIREQQEREQRLREQREREQQQQENQRREEEERE KQRLQEQREQEAQRELEKRYQLQREKLQREQLQREQLEREQREQEARDEWERNR EPAQREREQREWEQREREQREREQREREQREREQRELEREQQHEQREQQATE QQEQQEQQHREQEQREQEEREQQEREQQHREQMEGGHQPYEPNQPEEDQESSRRP YPSYEDYVRRDDEEEPANFEPDSEQSLYRNYSEEQEREQQQREQLYREQQERERQS LNPTDDEEDQQDRRQPYQTYEEYLRVNAEAEAAVAAAAQEQTDPENDPSAYENY EDYIRAQEERQEAEERERAVQEGWEDSRPEYPGYRQFVSNHHNSPQNAKQLRRQQ RPKL
D. virilis	MQLPTGKSCLLWQLLLLLVLGVCHAQFDEVHEDSERAAWLAGVFTPQPLKAAPW QAPTVNTPPPPPSGGLHVPGVGELFGLRGYKVIKNRPIDAYLGVRYAQVNGGLGRF QPARAVPFQARINATQSSANCAQFPEVQRLQVAEARGENVDDCLTLNIYAPAGAH DLPVLVFVHGEMLFDGGAEEGQPDYVLEHDVVLVSINYRLAPFGFLSALSDELPGN VALSDLKQALEWVQRNVRYFGGTPNQVTLVGQAGGATLAHALSLSPQAQHLFQQ LILQSGTALNPYLIDERPLETLDTFAQLARCPSAGRNLGPLYECLRRLRTSQLVSIFE QLLQQNEPRGLSALGGFKLVVGDRLGYLPEHPAALIASNSINNKTKPLIVGATKDA SAFILSRFYDQIQGLQSHNVSDYIKVVLRHTAQPQDHKIWHDWAIREIFTPEQVRFV NARSVTQGLLELSNLILYRAPVIYSIRLSHKKSPAYLYTFDYRGEHHRFGHLNSPLP FGVDASLSDDSVYLFPYPEEARNLNPEDKSLARALVAMWVNFAHSGIPNQNTNV WPKATSEYGPFLRFTNSRQSILELDQHFGEGINVPNLYAQYFNTTNITSGASATTTT TTTTTTTTRRPYVNFPGQGYRQPQQPEYRRPPVYNDYYARDQELRRQRLLQEQQE RDLKLREQQEQEQQQQEELMREEEERKREQQQREAEMQREPQQREQFEREQRQR AQQRREDEARQLVEQREQEERDKLVREQEVRQLQEREQQQLEQQRREQQERDQR LHEMPEPTEQQRQQQEREQEQRSQEELEQESREQREREQQQREQLQREHQDQGQP DLSDYPTYNDYVKAYTQWAQGQTFAGAEQEASAENPHENYPYAQPNENPQRLKL MRRHQRLRF
D. willistoni	MRWSSVKAQLEMLLATLLIFSICQCQARPNVYDDVHEENEKDAWLAGVYQPQPL RPAPWQRPDAARLPNPASDGGLQIPGVGELIGLPGYKFIGNRPIDAYLGIRYAQVGS

	GLGRFQQARSVPYQGQVNATKASPNCAQFPELQRIQDDEARGKNVDDCLTLDVY APSGASNLPVLVFVHGEMLFDGGSEEAAPDYFLEHDVVLVSINYRLAPFGFLSDLS SELPGNVALSDIHLALQWLQRNLRYFGGNPQQVTLVGQAGGATLVHALSLSGRAQ NLFQQLILQSGTALNPYLIDERPRETLATFARIARCPYAGQNLSPLYNCLSQLRTSEL LRAFQQLYEQNEPRGLTYLGGFKLVVGDSLGYLPQHPANLVVNNQNSSVPLIIGAT KDASAFILSRFYSEVQRVQSRNVSDYINVVLRHTAPPQHHKIWHDWALREIFTPDQ TPQFVNSQSISQHLLELSNLILYREPVIDSIRFTYKKVPTYLYIFDYRGQYHRFGHLN NPLPFGVDASLSDDSVYLFPYPVEASRLNPEDKSLARALVTMWVNFAQSGVPNPY NGVWPKASSEYGPFLRFTNSKESVLELDQHFGEGIYVPNLYGFDQILKSTTPQVSPT PTTTTTRTTTTTTTTRRPYAFNPYAQRPPQPSPQEPRRPSDQARYEQFLKAQERRRQ QYQREQQERLERERAQQQLREQHEREEQKRQEEERLKQREQEAQREREEHEQRLQ EERRRELQLQEERTREDQEKRLQQEREQEEREQHAQRPYETGPEEDSGYEPTPYPN YEDYVRDNRRQQEADTDPGTVPNTDNDIDNGNRPIGNDPSFDVNDPRSYSSYDAY VEAQMKLREEENRRYEEEQARYEQERQREEEERTRQQEEYGNHEPDQVHEEEEEQ RAHETYGQGQQQGEQENNLREEDNLPADFDPNDPRSYPSYDAYVEAQMKLREEE NRIYEQERVREEEERARQQEDHGHHFDEDDHLRHEEELVREEEERAEEQRNNQEE GEQDNNFREEEEEQERAREEDERARDEYELSREAEEEQAN
Rhodnius prolixus	MQAASSGPDIIQIEFITETTEGLLQGYALKARNGRTVIGFSGIPYAKPPIGALRFQDPE PPDKWDGIRQAQIYGNVCVQLINLIPAVRNIIIGNEDCLYLTVYTPKISSSSQYPVLV YFHGGAFAFGFGEIAHGPGFFMDHDVVLVTINYRLGSLGFLSMEDELLPGNFGLKD QNIALKWVQRNINNFGGNPNKVTIMGESAGGASVYYHMISPLSQGLYHGAISQSG TAFCPWAYSAPGLARKKAIKTADILGCPTDSTVDMVECLKIKHPYDIIGVHKNFIIW QNDPLTVFTPSPDPKAKNPFLPDEIHKLFPASVPWLTGCNNLEGLLKTAGFSVFRNS EKELEKQFSKLIPQILMYSDTSSNTHEVTNKLRRFYLDDNPVSSLTLRNFTEMVGD AWFVWPMIRSLKKHKGNNYVYYNTYIGKETLQSLFSSYRNVLNGTAHGDETLYL WDYSSVIEPHKGEDLKFSELLTKLWVNFATNGTPTPSGFEFAWPIWTAEHNKYVIF SNKGISEAEPFLEDQYFFLKRSGNVACGIQDPLPPKPWNGTLTKKVSKLCLQYLHPF YTKVIGSEDCLYLTVYTPKLNTTTLFPVTIYFHGGAFAFGAGQLEHGAQYMMDHN IVLVQVEYRLGAMGFLTMEDHVMSGNMGLKDQLMALKWVRNNIAQFHGDPNKV TLMGDSAGAASLFRGVISESGVSLAPWAITEPGVARRTAIKTAKSLKCPTDNTYEM LHCLQSVEVNNILSMYQRDDFLMNCGNPLFTFLPIIDSNSSTPFLPQNPFLMESAPVP WLTGCNSIDGFVLSGYFVSLPKWINFYGIIDIAFNQLMPKFLFYNSNKHATEISNKIR KFYFGNQEITSSSLKNLTNLFSDRFFFWPMMESLKLHDGPSYVYNFAYNSNLSYQF DHTSVRVLNGSDHEDETIFLWNVTHRFNQSFDKNDLELSEHLVKLFVDFVSYGTPT PAESEFKWPQWNNHDQRFISFGSNGTIIQDQKFYPHRMLFWSSLKQKELI

Figure 9. Glutactin Homologue Alignment

```
CLUSTAL 2.1 Multiple Sequence Alignments


Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: Dmel        1026 aa
Sequence 2: Dana        1164 aa
Sequence 3: Dgri        1310 aa
Sequence 4: Dpso        1057 aa
Sequence 5: Dvir         895 aa
Sequence 6: Dwil        1043 aa
Sequence 7: Rpro        1007 aa
```

---

clustalw.aln

```
CLUSTAL 2.1 multiple sequence alignment


Dmel            -----MKPLLLVLALCGAQVHAHSVGLR------PDYNDYSDEDTRRDWLPEPLKPVPWQ
Dana            -----MKPLYFVLALCVAQSCALYRGKR------GSDYDDDYSDEEESWQPDRVKKEPWK
Dgri            -------MQLLSLLLLLLVLGICHAQFD-----EIYS-----------YTPRPLRPAPWL
Dvir            MQLPTGKSCLLWQLLLLLVLGVCHAQFD-----EVHEDSERAAWLAGVFTPQPLKAAPWQ
Dpso            MQPLPSWVQLLALALCVCQICALQTRPR-----TDDIDYELSEENEGSFTPQPLRPPPWQ
Dwil            MRWSSVKAQLEMLLATLLIFSICQCQARPNVYDDVHEENEKDAWLAGVYQPQPLRPAPWQ
Rpro            --------------MQAASS----------------------------------------


Dmel            S-----------------ETRYAQPQEAVVQAPEVGQILGISGHKTIANRPVNAFLGIRY
Dana            VRHTPPVRVQQDISSQQAPQYAQDSRDVVVQAPEVGEIVGVRGYKSNANGLINAFLGIRY
Dgri            GE-----------VFKTTTPPPPAPPSGGVHVPGVGEVNGLPKYKVIKGRPIDAYLGIRY
Dvir            AP-----------TVN----TPPPPPSGGLHVPGVGELFGLRGYKVIKNRPIDAYLGVRY
Dpso            EAPQS----------------QEAPQEARVQVPGVGHIVGVRGYKTISNRPINAFLGVHY
Dwil            RP--------------DAARLPNPASDGGLQIPGVGELIGLPGYKFIGNRPIDAYLGIRY
Rpro            ----------------------GPDIIQIEFITETTEGLLQGYALKARNGRTVIGFSGIPY
                                     .  .    .  :      *   .  : .: *: *

Dmel            GTVGGGLARFQAAQPIG-YQGRVNATVQSPNCAQFPELDRLRLSESRGENVDDCLTLDIY
Dana            GKVGSGLSRFQAAQPTG-YQGRVNANVKSPNCAQFPELQRLRESEARGENVDDCLTLDIY
Dgri            AQVNSGLGRFQKARAVP-YAGRIDATQMSPNCAQFPEKQRLQEAEARGENADDCLTLNIY
Dvir            AQVNGGLGRFQPARAVP-FQARINATQSSANCAQFPEVQRLQVAEARGENVDDCLTLNIY
Dpso            AQVGGGLGRFQQARPVP-FQGRVDATTASPNCAQFPELQRLQTAEARGENVDDCLTLNIY
Dwil            AQVGSGLGRFQQARSVP-YQGQVNATKASPNCAQFPELQRIQDDEARGKNVDDCLTLDVY
Rpro            AKPPIGALRFQDPEPPDKWDGIRQAQIYGNVCVQLINLIPAVRNIIIGN--EDCLYLTVY
                .    *  *** ...   : .  :*   .  *.*: :          *:  :*** * :*

Dmel            APEGAN--QLPVLVFVHGEMLFDGGSEEAQ-PDYVLEKDVLLVSINYRLAPFGFLSALTD
Dana            APEHAS--NLPVLVFVHGEMLFDGGSEEAQ-PDYVLEKDVVLVSINYRLAPFGFLSALSD
Dgri            APAGPR--ELPVLVFVHGEMLFDGSAEEAQ-PDYVLEHDVVLVSINYRLAPFGFLSALSD
Dvir            APAGAH--DLPVLVFVHGEMLFDGGAEEGQ-PDYVLEHDVVLVSINYRLAPFGFLSALSD
Dpso            APEGAR--DLPVLVFVHGEMLFDGGAEEAQ-PDYVLENEVVLVSINYRLAPFGFLSALSE
Dwil            APSGAS--NLPVLVFVHGEMLFDGGSEEAA-PDYFLEHDVVLVSINYRLAPFGFLSDLSS
Rpro            TPKISSSSQYPVLVYFHGGAFAFGFGEIAHGPGFFMDHDVVLVTINYRLGSLGFLSMEDE
                :*  .   : ****:.**  :  * .* .  *.:.::::*:**:*****..:****   .

Dmel            ELPGNVALSDLQLALEWLQRNVVHFGGNAGQVTLVGQAGGATLAHALSLSGRAGNLFQQL
Dana            ELPGNVALSDLHLALEWLQKNLESFGGNPSQVTLLGQAGGATLAHALSLSGRATGLFQQL
Dgri            ELPGNVALSDLQLALEWVQRNIRYFGGASSRVTLIGQAGGATLAHALSLSPQAQHLFQQL
Dvir            ELPGNVALSDLKQALEWVQRNVRYFGGTPNQVTLVGQAGGATLAHALSLSPQAQHLFQQL
Dpso            QLPGNVALSDIHLALEWLQRNLPYFGGNSGKVTLVGQAGGASLVHALSLSGRATNLFQQL
Dwil            ELPGNVALSDIHLALQWLQRNLRYFGGNPQQVTLVGQAGGATLVHALSLSGRAQNLFQQL
Rpro            LLPGNFGLKDQNIALKWVQRNINNFGGNPNKVTIMGESAGGASVYYHMISPLSQGLYHGA
                 ****..*.* : **:*:*:*:*  *** . :**::*::.*.: .:   :*  :  *::

Dmel            ILQSGTALNPYLIDNQPLDTLSTFARLARCPPPSINPSAQGLKPLYDCLARLPTSQLVAA
Dana            VLQSGTALNPYLIDERPLDTLDTFASFAGCPTAAN---TRSLAPLYDCLGSLPTSNLVAA
Dgri            ILQSGTALNPYLIDERPLETLATFASLARCPTAGRN-----LAPLYQCLNRMRTSQLVEI
```

```
Dvir            ILQSGTALNPYLIDERPLETLDTFAQLARCPSAGRN-----LGPLYECLRRLRTSQLVSI
Dpso            ILQSGTALNPYLIDERPLETLDTFARLARCPNAGRS-----LAPLYDCLSRLPTSQVVSA
Dwil            ILQSGTALNPYLIDERPRETLATFARIARCPYAGQN-----LSPLYNCLSQLRTSELLRA
Rpro            ISQSGTAFCPWAYSAPGLARKKAIKTADILG--------------------CPTDSTVDM
                : *****: *:  .        ::                            *.. :

Dmel            FEQLLLQNEHLGLTQLGGFKLVVGDPLGFLPSHPASLATN---SSLALPMIIGATKDAS-
Dana            FEQLFNRNEAIGLSSLGGFKLVVGDPLGYLPKPPVALAAN---ASNALPMIVGATKDAS-
Dgri            FEQLFRNNEPLGLSALGGFKLVVGDRLGYLPNHPAALVAS---NNSTKPLIVGAAKDAS-
Dvir            FEQLLQQNEPRGLSALGGFKLVVGDRLGYLPEHPAALIASNSINNKTKPLIVGATKDAS-
Dpso            FEQLFQQNEARGLNFLGGFKLVVGDPLGYLPEHPAALAAN--SANSTVPMIVGAAKDAS-
Dwil            FQQLYEQNEPRGLTYLGGFKLVVGDSLGYLPQHPANLVVN--NQNSSVPLIIGATKDAS-
Rpro            VECLKIKHPYDIIGVHKNFIIWQNDPLTVFTPSPDPKAKNPFLPDEIHKLFPASVPWLTG
                .: *  .:    :    .* :  .* *  :.  *     .    .    :: .:.   :

Dmel            AFIVSRIYDQLARLQSRNVSDYIDVVLRHTAPPSEHRLWKQWALREIFTPIQE-QTASLQ
Dana            AFILSRFYNQIEHVRSYNVSDYIDVVLRHLAPPSEHQIWKQWALREIFSPDQV-RTVNSI
Dgri            AFILSGFYDQIQNLNSRNLSDYINVILRHTAPPQHHKVWHDWALQEIFTPEQK-RYVNTN
Dvir            AFILSRFYDQIQGLQSHNVSDYIKVVLRHTAQPQDHKIWHDWAIREIFTPEQV-RFVNAR
Dpso            AFILSRFYDQIQRVQSQNVSDFINVVLRHTAPPSQHQLWLDWARREIFSQEHD-RTASAY
Dwil            AFILSRFYSEVQRVQSRNVSDYINVVLRHTAPPQHHKIWHDWALREIFTPDQTPQFVNSQ
Rpro            CNNLEGLLKTAGFSVFRNSEKELEKQFSKLIPQILMYSDTSSNTHEVTNKLRRFYLDDNP
                .  :. : .       * .. :.  : :           .   :*: .  :     .

Dmel            TVAPGLLELSNYILYRAPVINSIS-QSYRSVPAYLYTFDYRGEHHRFGHLSNPLPFGVDA
Dana            SVSQGLLELSNMILYRAPVIYSIR-NSYRNAPVYLYNFAYRGEHHRFEHVKSPLPFAVDA
Dgri            SVAQGLLELSNLIMYRSPVIDTIR-LSHQKTPAYLYTFDYRGQYHRFADVKNPLPFGVDA
Dvir            SVTQGLLELSNLILYRAPVIYSIR-LSHKKSPAYLYTFDYRGEHHRFGHLNSPLPFGVDA
Dpso            SVSQGLLELSNMILYRAPVVYSIR-LSHKKSPVYLYTFDYRGEYHRFGHLDNPLPFGVDA
Dwil            SISQHLLELSNLILYREPVIDSIR-FTYKKVPTYLYIFDYRGQYHRFGHLNNPLPFGVDA
Rpro            VSSLTLRNFTEMVGDAWFVWPMIRSLKKHKGNNYVYYNTYIGKETLQSLFSSYRNVLNGT
                  :  * :::: :     *   *   . :.   *:*   * *:      ...   .  .:

Dmel            SLSDDSVYLFPYPPEASRLNPLDRSLSRALVTMWVNFATTGVPNPS--SGVWPQATSEYG
Dana            SLSDDNVYVFPYPVETSQLNAVDRSLSRALVTMWVNFATTGVPNPN--FGVWPRASSEYG
Dgri            SLSDDSVYLFPYPEETSKLNPDDMSIARALVAMWVSFAQTGIPNPN--PNVWPKATSEYG
Dvir            SLSDDSVYLFPYPEEARNLNPEDKSLARALVAMWVNFAHSGIPNQN--TNVWPKATSEYG
Dpso            SLSDDSVYLFPYPPEASRLNPVDKSLSRALVTMWVNFAVTGIPNQN--PGVWPKATSEYG
Dwil            SLSDDSVYLFPYPVEASRLNPEDKSLARALVTMWVNFAQSGVPNPY--NGVWPKASSEYG
Rpro            AHGDETLYLWDYSSVIEPHKGEDLKFSELLTKLWVNFATNGTPTPSGFEFAWPIWTAEHN
                : .*:.:*:: *.      :  * .::. *. :**.** .* *.      .**  ::*:.

Dmel            PFLRFTNNQQSPLELDPHFGEGIYLP-NYRVIYK----PTTNFSPPITTTTTTTTTTTT
Dana            PVLRFSNSRENMLELDPHFGEGINIPNLYSNYFN----TTSKTTPAVIPVFTTTTTTTTT
Dgri            PFLRFTNSKQNILELDQHFGEGINIPNLY---------PMYFTTSRTSSGTTTTTRRPYS
Dvir            PFLRFTNSRQSILELDQHFGEGINVPNLY---------AQYFNTTNITSGASATTTTTTT
Dpso            PFLRFTNSRQNMLELDPHFGDGLYAPNLY---------GQYFNTTNSTTTATTTTTTTTT
Dwil            PFLRFTNSKESVLELDQHFGEGIYVPNLYGFDQILKSTTPQVSPTPTTTTTRTTTTTTTT
Rpro            KYVIFSN--KGISEAEPFLEDQYFFLKRSGNVACGIQDPLPPKPWNGTLTKKVSKLCLQY
                  : *:*  :.  * : .: :                      .        .:.

Dmel            SRPYAYNPYANWQNRPSQQHPNWHPADPEYVRAQEARQQEFIRE----------------
Dana            TRPYVHNPYNNWNRQIQQTGPSRDAAEEERRKQEFREYQRRQEEERRREYQLRLQQRQEQ
Dgri            GLPPAYPEYNPRTQEAHWK-----------QQMREQEEQLARENEKRRLEQQQREEQLQQ
Dvir            TTTTTRRPYVNFPGQGYRQ-----------PQQPEYRRPPVYNDYYARDQELRRQRLLQE
Dpso            RRPYVYNPYPSWQQSPYGYNTTTTTSTTTTTTRRPTYAYNPYANWQQNRQPQPPQAFPRR
Dwil            RRPYAFNPYAQRPPQPSPQEPR--------RPSDQARYEQFLKAQERRRQQYQREQQERL
Rpro            LHPFYTKVIGSEDCLYLTVYTPKLNTTTLFPVTIYFHGGAFAFGAGQLEHGAQYMMDHNI
                  .

Dmel            --REQRRREQQLRDQQRYPQQEPREQQDERIRQQREQEERLRQQREQEERLRQQRELEER
Dana            LQRESLEREKKLQEQREQQQREQELQEQQREQQQREQQRREQQWREQQQREQQWREQQER
Dgri            ENAQREQLERDQRQRWAAEELALKEYELRQQREREQQQREQHEREQPDATQEQLDP--EQ
Dvir            QQE--------------------RDLKLREQQEQEQQQQEELMREEEERKREQQQR--EA
Dpso            PTDPAYERAQEIRRQQVIREQQEREQRLREQREREQQQQENQRREEEEREKQRLQEQREQ
Dwil            ERERAQQQLREQHEREEQKRQEEERLKQREQEAQREREEHEQRLQEERRRELQLQEERTR
Rpro            VLVQVEYRLGAMGFLTMEDHVMSGNMGLKDQLMALKWVRNNIAQFHGDPNKVTLMGDSAG
                                      .     :  ..     .

Dmel            ---------------------------------IRQQQEREQYEREQQEREQREREELER
Dana            EQQERQYGDVNRDPYTVTYDAYGREITPNSAEQERQKQEREQLEREEQQREQQRREYEER
```

```
Dgri            EQHSEEQPDQESRWQQEREQQLREQ-------QRREQQQREQRQREQQQREQQQREQQHR
Dvir            EMQ-----------------------------REPQQREQFEREQRQRAQQRREDEAR
Dpso            E-------------------------------AQRELEKRYQLQREKLQREQLQREQLER
Dwil            EDQ-----------------------------EKRLQQEREQEEREQHAQRPYETGPEED
Rpro            AASLFRGVISESGVSLAPWAITEPG-------VARRTAIKTAKSLKCPTDNTYEMLHCLQ
                                               *    :   . :      .

Dmel            QQRERERQQQPEQQPEYNPEPVNP------------------WGYPVQEPQPDDNPEDGRL
Dana            LEQDRQQRERERQQEREQQERVQP------------------EGQ----YQPEENTEDGRQ
Dgri            QQPEREQEQPAQQQPHYSQSDVPPEPNDFPTYEDYVKAYEQWVRAAQEQEQHQQSQEERE
Dvir            QLVE----------------------------------------------QREQEERDKLVRE
Dpso            EQREQEARDEWERN-------------------------------REPAQRERERQREWE
Dwil            SGYEPTPYPNYEDYVRDNRRQQEADTDPGTVPNTDNDIDNGNRPIGNDPSFDVNDPRSYS
Rpro            SVEVNNILSMYQRDDFLMNCGNP--------------------LFTFLPIIDSNSSTPFL
                                                                        .

Dmel            PYPSYEQYGP-EGNENLPETDANRNFSEEDREQQQQEQLRREQQ----------------
Dana            AYPSYEDYLKSQQEEQSPESDPESNYAEEQQREQERLEQERQLR----------------
Dgri            REAREQEELEQQQREQQDRQQQERPHHEEEQPEATDEQPEPEQEQHSQTVLDQELRWKQV
Dvir            QEVRQLQEREQQQLEQQRREQQERDQRLHEMPEPTEQQRQQQER----------------
Dpso            QRERERQRERERQRERERQRERERQLEREQQHEQREQQATEQQ----------------
Dwil            SYDAYVEAQMKLREEENRRYEEEQARYEQERQREEEERTRQQEE----------------
Rpro            PQNPFLMESAPVPWLTGCNSIDGFVLSGYFVSLPKWINFYGIID----------------
                              .                 .

Dmel            ----EQQEREYQLQLEREQQEREQQERGQQEPGPEE----------------------YPS
Dana            ----EREHREYEERLEQERQQRERERQQEREEQGGQHQPEENPED---------GRDAYPT
Dgri            LEQDERELRQREEREYEQREQQRRDQQQREQQGREHQEPAQQEPHSSQSDAPPEPNNFPT
Dvir            ----EQEQRSQEELEQESREQRERERQQQREQLQREHQDQGQ-----------PDLSDYPT
Dpso            -EQQEQQHREQEQREQEEREQQEREQQHREQMEGGHQPYEPNQP---EEDQESSRRPYPS
Dwil            -----YGNHEPDQVHEEEEEQRAHETYGQGQQQGEQENNLREEDNLPADFDPNDPRSYPS
Rpro            --IAFNQLMPKFLFYNSNKHATEISNKIRKFYFGNQEITSSSLKN-LTNLFSDRFFFWPM
                               ...    .   :      .                     :*

Dmel            YEEYSRALQ---------------EKNAERDRIYAEEQERERQQQETLLQ----------
Dana            YEDYVRAQQGEQSPESDPDRNYAQEQQREQERLEQERQLREQEHRQYEER----------
Dgri            YEDYVKAYE---------------QWAGAAQEQEQQQQSQEDNEQEARDQQEREQQLRQI
Dvir            YNDYVKAYT---------------QWAQGQTFAGAEQEASAENPHEN-------------
Dpso            YEDYVRRDD---------------EEEPANFEPDSEQSLYRNYSEEQ-------------
Dwil            YDAYVEAQM---------------KLREEENRIYEQERVREEEERAR-------------
Rpro            MESLKLHDG---------------PSYVYNFAYNSNLSYQFDHTSVRVLN---------
                 :                              ..

Dmel            ------------------------------------------------------------
Dana            ------------------------------------------------------------
Dgri            EQEELAKHASEPEGKPYRYDYSSDTSYMLAVAEWEAAGPDYKPYQRTQEEIEEDARYERK
Dvir            ------------------------------------------------------------
Dpso            ------------------------------------------------------------
Dwil            ------------------------------------------------------------
Rpro            ------------------------------------------------------------


Dmel            ------------------------------------------------------------
Dana            ------------------------------------------------------------
Dgri            RKILMNMRKDYEEGVQHPREPGAEPRTHEFNNFESFIAAHSKWEEEHPVQKEELESPEQR
Dvir            ------------------------------------------------------------
Dpso            ------------------------------------------------------------
Dwil            ------------------------------------------------------------
Rpro            ------------------------------------------------------------


Dmel            ------------------------------------------------------------ENQ
Dana            ------------------------------------------------------------LQR
Dgri            AQRDRETRGRVEQERPQHQVLDLELTQPQNGYDNFNDYVNGRALLTAARQDRDQQQQSQE
Dvir            ------------------------------------------------------------
Dpso            ------------------------------------------------------------EREQQ
Dwil            ------------------------------------------------------------
Rpro            ------------------------------------------------------------


Dmel            QHPEQSLPEEQP---------------------THPNYEAYDGDRSYAEEQEREQQRRDQ
```

```
Dana            EREQQEREQEQPQGQPDNDLSRYPSYDEYLRSQQEGRPSDSDSDPNYDEEQAREQQRRDQ
Dgri            EREQQAREQYIREQQQRHQLEREQEEQERAQYEPHDDYFLNASPWPEAGQEQDQQQQSQE
Dvir            ------------------------------------------------------------
Dpso            QREQLYREQQERERQSLNPTDDEEDQQDRRQPYQTYEEYLRVNAEAEAAVAAAAQEQTDP
Dwil            ------------------------------------------------------QQEDHGH
Rpro            ----------------------------------GSDHEDETIFLWNVTHRFNQSFDKN


Dmel            VEQEREE--QPDEDQGEEYERSPDEEEAAEQDVLKVEDFPSYEAYLEAATKLREEQEEQE
Dana            LRQEQEQEYGPSGGYPEESDDQDQEYQPNVDPKNDPSNYATYEDYVKAQLKLREEEEERD
Dgri            EYDQEAREQQEREQRERELEREEQQRSDAQPDTDTYGLDDSDLNADLWPQASEEEEADR
Dvir            ------------------------------------------------------------
Dpso            ENDPSAYENYEDYIRAQEERQEAEERERAVQEG----------------WEDSRPEYPGY
Dwil            HFDEDDHLRHEEELVREEEERAEEQRNNQEEGEQD---------------NNFREEEEEQ
Rpro            DLELSEHLVKLFVDFVSYGTPTPAESEFKWPQWNN--------------HDQRFISFGSN


Dmel            ---KLEEERYRAQQEEEDRIQAERERNSRN----------------------
Dana            RALREEEERQFAQAEAEQEAARQSQPTSIVFRRFLNAPRHIKVIHRQQGRRH
Dgri            YPEYSNGHNYPQRVKLMRRHPRLRS----------------------------
Dvir            -YPYAQPNENPQRLKLMRRHQRLRF----------------------------
Dpso            RQFVSNHHNSPQNAKQLRRQQRPKL----------------------------
Dwil            ERAREEDERARDEYELSREAEEEQAN----------------------------
Rpro            GTIIQDQKFYPHRMLFWSSLKQKELI----------------------------
                     : .    .
```

Figure 9. Glutactin Homologue Alignment. There are no yellow, highlighted regions to indicate transmembrane domains as Glutactin is a secreted protein.

Figure 10. Glutactin Phylogenetic Tree

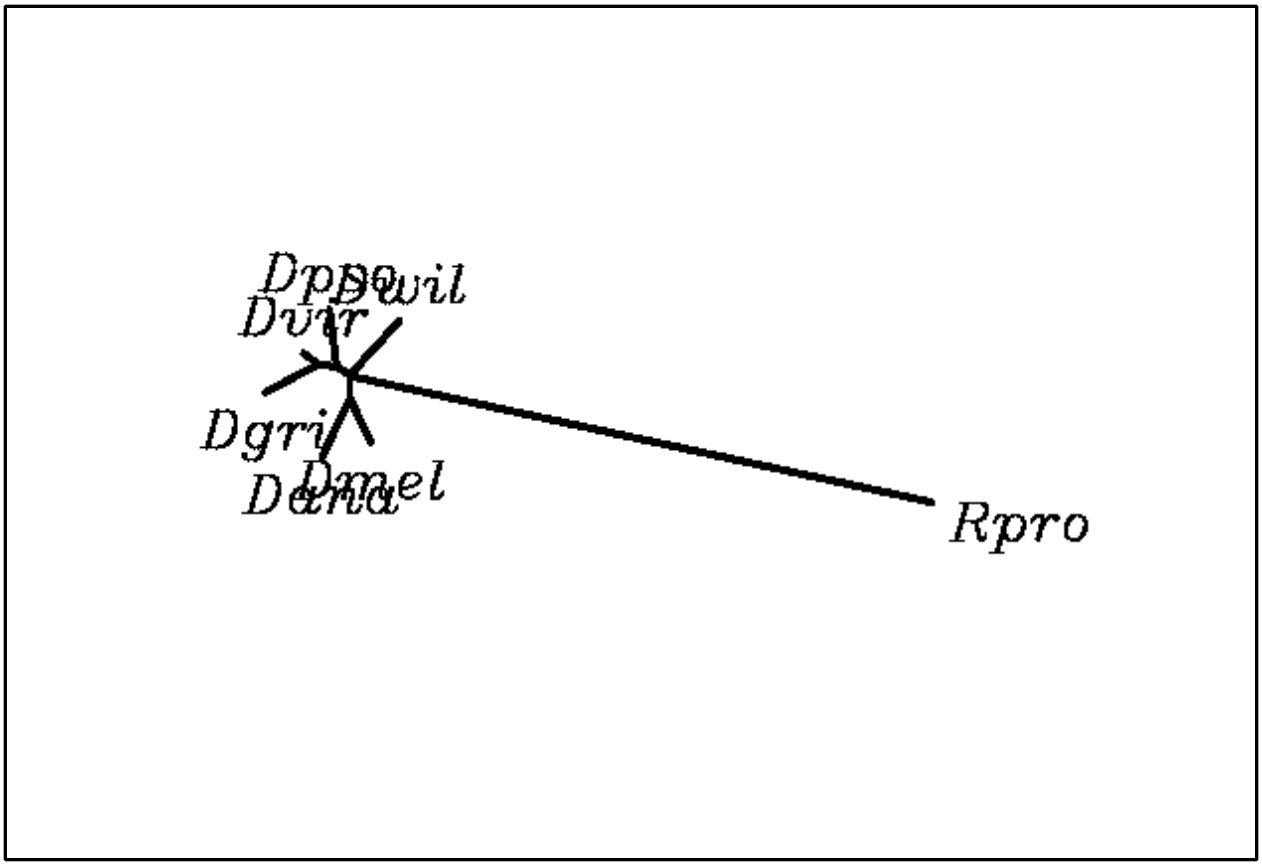


Table 5. Neurotactin Homologues

D. melanogaster	MGELEEKETPPTETTAAQQEALEEPKETDKMLDKKEDAKEKTPSPQTSKPASPNA GKKSSPVAEKKIDDAELAKSKSGNGEEIIDIPAENGTKPDSADDKKISKEEREVKP KKIPIGGLKLPGFFMKNKPKADGDGAEGELLEKEKEEDKDKEANGDAATGSGKD EQKSRPGLGERLRSFFARKPSAEKEKKQLVNGDADAKSEATAEATPAEDASDAPP KRGLLNAIKLPIANMIPKKKSNDDVELGLGKAGLASMETLDDSLKDQDTVDRAP VKTNGTEELKGELKDEKLAAEEKLAAEEEEQNRPVSLLTRLRGYKCSVDDALIVF GILLFVLLLGVIGYVLTHETLTSPPLREGRYIMAVTGCGPVEGVKEDGAFAFRGIP YAKPPVDRLRWKPAELIDDINMCWNDTLQTHNSSVVCTQRLGNGTTVGDEDCLY LDVVTPHVRYNNPLPVVVLIGAESLAGPSPGILRPSARYSRSHDVIFVRPNFRLGVF GFLALDALTKEAHPPTSGNYALTDIIAVLNWIKLNIVHFGGDPQSVTLLGHRAGAT LVTLLVNSQKVKGLYTRAWASSGSAILPGKPLSESGKQNEQLMATLECADIQCLR EASSERLWAATPDTWLHFPVDLPQPQEANASGSRHEWLVLDGDVVFEHPSDTWK REQANDKPVLVMGATAHEAHTEKLRELHANWTREEVRAYLENSQIGALGLTDE VIEKYNASSYASLVSIISDIRSVCPLLTNARQQPSVPFYVVTQGEGPDQLATVDAD VQAILGRYEPHTVEQRRFVSAMQQLFYYYVSHGTVQSFVQNRRVINVGQDAQPE EDYLPCNYWISKDIVPRYARVD
D. ananassae	MGELEEKETPPTETTTSQQEALDEPKETDKMLDKKEDDAKEKTPSPQTSKPASPN AGKKSSPLAEKKAVEGDTQKPGNGEEIIDIPAENGTKSSEGDDKKISKEEREVKPK KIPIGGLKLPGFFMKNKPKAEGDGAEGELLEKEKEEDKDKEANGDAATTSGKDE QKARPGLGERLRSFFVRKPAAEKEKKQLVNGDADAKSEATAEAAPAEDAAEAPP KRGLLNAIKLPIANMIPKKKNNDDVELGLGKAGLASMETLDDSLKDQDTVDRAP VKTNGTDEIKSDLKDEKLAAEQKLAAEEEERERPVPLLTRLRGYKCSVDDALIVF GILLFVLLLAVIGYVLTHETLTSPPLREGRYIVAVTSCGPVEGVKEDGAFAFRGIPY AKPPVDGLRWRPAELVDSIDGCWNDTLQTHNSSVVCTQRLGNGTTVGEEDCLYL DVVTPHVRYTNPLPVVVLIGAESLAGPSPGILRPSARYSRSHDVIFVRPNFRLGVFG FLALDALTKETRPPTSGNYALTDIIAVLKWIKLNIANFGGDPKSVTLLGHRAGATL VTALVTSNKVKDLYTRAWASSGSAILPGRPLSVSENLNQQLMGTLECADVECLRE ASSERLWAATPDTWLHFPVDIPQPQEFNAGQRHEWLVLDGDVLYQHPSESWKQD KATDKPLLVMGATAHEAHSERLRELHANWTREEVRAYLTNSQIGAMGLTEEVLE RYNASSYASLVSIITDIRTVCPLLTNARLQLGVPFYVVTQGEGPDQLATVDADVQ AILGRYEPHTVEQRRFVSAMQQLFYYYVSHGTVQTFEPQRRVINVGQDAQSQEDF EPCNYWISKDIVPRYARVD
D. grimshawi	MGELEEKETPQTETTTVQESTLDEPKETDKMLDAKKEKTPSPQTSKPGTPNTGKK SPEVPLEVLSKNVVEKPLNNGEEIIDIPTENGNKPGAAIDVTDEKKISKEEREVKPK KIPIGGLKLPGFFMKNKPKGEGDGAEGELLEKDNEQEQEEKEANGGDATTTPNKD EQKPRTNLGQRLRSFFVRKPAADKEQQKKQLNNGDADAKSEATAEAAAVEDAD PPPKRGLLNAIKLPIANMMPRKKAGDDVELGLGKAGLASMETLDDSLKDQDTVD RAPLKTNGADELKSDLKDDKLAAEQKLAAEEEERDRPISLWCRLLGYKCSVDDA LIVFGILLFMLLLAVIGYVLTHETLTSPPLREGRYISAVTSCGMVEGVKEDGAFAFR GIPFAKAPIDDLRWRPAQLIEKIDDCWNGTLQTHNESTVCTQRLGNGTTIGDEDCL YLDVVTPHVRYTNPLPVVVLIGAESLTGPSPGILRPSARYSRSHDVIFVRPNFRLGV FGFLALDALTKESHPRTSGNYALTDIVAALKWIQRNIMHFGGDPKSVTLLGHRAG ATLVSTLVSSAKLEEYYTRAWASSGSAIFPGKPLHESERSNQQLMTTLECQDAVC LRETSSEKLWAATPDTWFHFPTTSPQELELSNANRHEWLVLDGDILQQHPSESWK QVPTASQLQHPRYVMGATAHEAHTLKLRERHANWTAEEVRSYLQASKLGELGL TDEIIQRYNATNYASLVAIITDIQTVCPLLTNARLQPSVPFYVVTQGEGDDQLATV DADVQAILGRYEPHTVEQRRFVSAMQQLFYYYVSHGTVQTFESRRRVIIVGQDAQ SQEDLEHCNYWISKDIVPRYARVD
D. pseudoobscura	MGELEEQETPPTETTAAQQETLEEPKETDKMLDSKKTEKTPSPQSSKPATPNAGK

	KSPVGEKTNEEKGHEIIDIPEEKEKPQPTEGSGAGEEKKISKEEREVKPKKIPIGGLK LPGFFMKNKPKAEGDGAEGELLEKDAEDKDKEATAVNGDAAAAAAAAEPKPRP GLGQRLRSFFVRKPGAEKEKKALVNGDADAKSEATAEAAAAEDAAAVDPPPKR GLLNAIKLPIANMIPKQRKNGDDVEMGLGGKAGLASMETLDDSLKDQDTVDRAP VKNANGTDDIKSELKDEKLAAEQKLAAEEDERDRPISLLQRLRGYKCSVDDALIA FGILLFVLLLAVIGYVLTHETLTSPPLREGRYILAVTGCGPVEGVKEDGAFAFRGIP YAKAPVDELRWRPAQLIDSIDGCWNDTLQTHNSSVVCTQRLGNGTTVGDEDCLF LDVVTPHVRYTNPLPVVVLIGAETLAGPSPGILRPSARYSRSHDVIFVRPNFRLGIF GFLALESLTKEAHPRTSGNYALTDIIAVLKWIQLNIVNFGGDPKSVTLLGHRAGAT LVSALVSSPKLKDLYTRAWSSSGSAIFPGKPLAEAEKLNQQLMVTLDCSDVGCLR DASSEKLWAATPDTWLHFPQDLPQPQEINSGSRHEWLVLDGDVLFQHPSESWKR DQATEQPLLVVGATAHEAHTEKLRELHANWTKEKVRDYLNNSQIGALGLTDEVI ERYNATNYAALVSIITDIRTVCPLLTNARLQPSVPFYVVTQGEGEDQLATVDADV QAILGRYEPHTVEQRRYVSAMQQLFYYYVSHGTVQSFEPRRRVINVGQDAQPQE DYEHCNYWISKDIVPRYARID
D. virilis	MGEIVEKETPQTEATAVQEPLEEPKETDKMLDTKKDDAKEKTASPQTSKPASPNA GQKSPTLPQADPAEKPAGGNGEEIIDIPTENGNKPNGAADEKKVSKEEREVKPKKI PIGGLKLPGYFMKNKPKAEGDGAEGELLEKDKELAEEKEANGDASKEEQKPRTG LGQRLRSLFVRKPAADKEQQKKPLTNGDADAKSEATAEAAPGEDADPPPKRGLL NAIKLPIANIMPKKRGGDDVEMGLGKAGLASMETLDDSLKDQDTVDRAPLKSNG ADELKSELKSDKLAAQQKLAAEEEERERPVSVWCRLRGYKCSVDDALIVFGILLF MLLLAVIGYVLTHETLTSPPLREGRYISAVTSCGMVEGVKEDGAFAFRGIPYAKAP LENLRWRPAQLISKIDDCWNGTLQTHNGSAVCMQRLGNGTTIGDEDCLYLDVVT PHVRYTNPLPVVVLIGAESLAGPSPGILRPSARYSRSHDVIFVRPNFRLGVFGFLAL DALTKESHPRTSGNYALTDIIAVLKWIQLNIVHFGGDPKSVTLLGHRAGATLVSVL VSSAKLEGLYTRAWASSASAIYPGKPLQESERLNQQLMATLECQDAECLRAASSE KLWAATPDTWFQFPATVPQAEELSNANRHEWLVLDGDILYQHPSESWKQVPTAS QLHQPRLVMGATAHEAHTLKLRELHANWTAEEVRSYLQNSQLGARGLTDEIIQR YNATSYAALVAIITDIRTVCPLLTNARLQPTVPFYVVTQGEGDDQLATVDADVQA ILGRYEPHTVEQRRFVSAMQQLFYYYVSHGTVQSFDTRRRVIIVGQDAQAQEDHD HCNYWISKDIVPRYARVD
D. willistoni	MGELEEKETPPTETTTNASSGHQQETLDEPKETDKMLDKKDETTSQEKTATPHNS KPSTPDAKKKLENDEKKSNGGEEIIDIPNENGTKPSENEDKKISKEEREVKPKKIPI GGLKLPGFFMKNKPKSEGDGAEGELLEKEKDDDENSKEAKVENGEAKGKDDQQ QQQQQKSRPGLGERLRNLFVRKPATDKEKKQQLINGKETADNDNKSEATAEANG QEESSVDPPPKRGLLNAIKLPIANMIPKKKSPDDVELGLGGGNKAGLASMETLDD SLKDQDTVDRAPVKSNGTDDLKSIELKDDKLPANEKLAAEEEEQERPASLWTRLR GYKCSVDDALIVFGILLFVLLMAVIGYVLTHETLTSPPLREGKFILAVTGCGPVEG VKEDGAFAFRGIPYAKAPIEERRWRPAQLIDSIDGCWNDTLQTHNSSVVCTQRLG NGTTVGDEDCLYLDVVTPHVRYTNPLPVVVLIGAESLAGPSPGILRPSARYSRSHD VIFVRPNFRLGIFGFLALDALTKESHPRTSGNYALTDIIAVLKWVQLNIVNFGGDP KSVTLLGHRAGATLVTALVSSPKLQGLYTRAWASSASAIFPGKPLSESEKVNQQL MATLDCQDVECLREASTEKLWAATPDTWLHFPMDLPQRQELSSSGQRHEWLVL DGDILYQHPSESWKLEQATDKPLLVMGATAHEAHSLKLRELHANWTKAEIQDYL NKSQLGAKSLTDEVMQLYNATNYASLLAIITDIRTVCPLLANARLQPSVPFYVVT QGEGDDQLANVDADVQAILGRYEPHTVEQRRFVSAMQQLFYYYVSHGTVQSYD SRRRVINVGQDAQPQEDHKNCNYWISKDIVPRYARID
Ceratitis capitata	MGELEEKETGPSETTPQQEVVDEPKETDKMLDKKEDTEAVEQQTENAGVKDKNP SPNTEKKRAASSKPATPTSLGGKKSSIQSVEKLATNGVDDAQSVKANGGSGEEIID IPESTKTEEGDEKKISTEDREVKPKKIPIGGLKLPGFFMKNKPKADGDGADGELLE RENKDEMPAETVNGESKSSKKDEKPRSSFGERLRNFFVRKPTADKQQNKQTANG DADAKSEATAEAAAADGETNNTTNASDAPPQKRGLLNAIKLPIANMIPRKKTDD DVELGMGKAGLASMETLDDSLKDQDCVDKAPARNNGTEDPAKLKKATSSEIKQ

	ESPEDKATEEPEAPMSFADRLRSYRCSVDDGLIVLGILLFLALIAVIGYVLSLETLTS PPVREGRFIDAVTGCGMVEGLQEDGAFAFRGIPYALPPLGDQRWRPAQPIGGIDD CWNGTLKAHNVSNYCTQRLGNGTIVGDEDCLYLDIVTPHVRYNSPVPVIVLIGAE TLTGPSPGVLRPSARYSRSHDVIFVRPNFRLGAFGFLALEALTKDAYPRTSGNYAL SDIIAALKWIQLNIAHFGGDPKSVTLLGHRAGGTLVSALVTSNKVEGLYKNAWVS SSSAIFPGKPLEESEKRNAQFMTALDCSSISCLRNAATEQIWAATPDTWLHFPVDM PTTDENPSTRHEWLVLDGNILQRHPADAWKEEHTGKPKLVMGTTAHEAHTEQLR QRHANWTSEEIRAFIATSKIGALNLTDEAIRLYNATNYRSLVAMITDIRTICPLLTN ARLQPTVPFYVVQQGEGENQLATVDTDVLAILGRYEPHTAEQRRFMSSMQQLFY YYISHGTVPQYDPSRRVLSIAQDPLPQSDHPNCNFWISHDIVPRYARVD
Musca domestica	MGEQEEKEIPQSENSTQPEVAEAEPKETDKMLEKKDDKPSAEDTKSKNGSKPTTP TATEKTEKKTPAGSPTSEKKKEEIIDIPEEENNDEATKTNGAGGEGDERKMSGEER EVKPKKIPIGGLKLPGFFVKNKPKGEGDGAEGELLEKEPKEEEAKTAEEKTSKKEE KPGKNFGQRLRSFFVRKPAAEKQAKQAASNEADNKSGKFFPSLFSPTFNCKRGLL NAIKLPIANMVPKKKSDDDVELGMGKAGLASMETLDDSLKDQDVVDKAAVTKN GNDDVGKLKKSPSGEIKQASSDEKSPDGDEEQPISFCQRLRSYKCSVDDALIALGI LLFVLLIAVIGYVLSHEGQTSPPIREGRFMDAVTGCGMVEGLKEDGAFAFRGIPYA VPPVGDRRWRPAQTIESINDCWNGTLKAHNTSEVCTQILGNGTVVGDEDCLYLD VVTPHVRYENPLPVVVMIGAETLMGPSPGIVRPTARFSRSHDVIFVRPNFRLGPFG FLALEMLTKDTYPHTSGNYALSDIIAALNWIHLNIRHFGGDPKSVTLLGHRAGATL VSALITSKKAEGLFTRAWISSASAILPGKPLSESEKKNEELAQALDCKDVACMRNV EAEKLWDAIPDTWLHFPADIPAAEENTTAHHEWLVLDGDILKEHPAESWKKEHV GKPLLVMGTTAHESHTESLHKRFGNWTAEEIRSFLENSKIGALNLTDEAIQRYNAT NYRALVAMITDIRTVCPLLINARLQPSVPFYVVSQGEGEEQLATVDADIQAILGRY EPHTVEQRRFVSSMQQLFYYYVAHGNVPSYDPRRRVIDVGQDPLSQEDHANCNF WISNDIVPRYARVD
Glossina morsitans	MGELEEKETPATESTTQHESAEEPKETDKMLEKKEESIEPHSEKSADNKKKSNSKP TTPVSERKTSEEMRSNTSEEIIDVPVESDAKTTIDEENHDNGDDKKISTEDREVKPK KIPIGGLKLPGFFMKSKPRVDGDGADGELLEKEHKEETTSEVKLAKKDEKPGKNF GKRLRDFFVRKPAAEKQAKQQAASTGDADTKSEATAEAGPASEGNMRTNSDGPP PKRGLLNAIKLPIANMIPKKSPDDDVELGMGKAGLASMETLDDSLKDQDCVDKA SPKNNGVDEISKLKKTPSGEIKQASEEKTFSDAEDVQISFCQRLRSYKCSVDDALIA LGILLFIFLIAVIGYVLSYETLTSPPIREGRFMDAVTGCGMVEGLKEDGAFAFRGIP YALPPTGDRRWRSAQTIEGIEDCWNGTLQAHNSSAMCTQMLGNGTVIGDEDCLY LDVITPHIRYDSPLPVIVMISAESLTGPSPGVLRPSARYSRSHDVIYVRPNFRLGAFG FLAVEVLTKDSYPHTSGNYALSDIIAALKWIKLNIVHFGGDPKSVTILGHRAGGTL VTALITSKKVEGLYTRAWISSASAIYPGKSLVESEKRNQEFLQALECQDVKCLRAA SAEKIWDATPDTWLHFPVDIPTAVENKTAHHEWLVLDGDILQEHPAESWKKEHA GDPVLVMGTTAHESHTEKLRRRFANWTAEEIRNFIESSKIGTLNLTDEAIRRYNAT NYPALVSMITDIRTICPLLINARLQPSAPFYVVTQGAGEEQLATVDADVQAIMGRY EPHTVEQRRFVSSMQQLFYYYVSHGTVPQYDPRRRVINVGQDPLSQSDYPNCNF WISNDIVPRYARID
Aedes aegypti	MGETDDKDTAPPTAPTEKPTTETPAAPEPHHEPEADPKETEKLLATREEKTDDKE TEKPADEKKRSSENVNGTEEIINIPDETSTETVKVEDEKAKTDKPNKVQAEEREVK PKKVPAGAFKLPGFFNKGKAKEADGADNELLEKQETDKDAKAAEAAEEKPKRP GFFANLRLRNPFAKKQAEPVPTEETKAKPEEEENDEVTAEATDKKPEDESAVEKK EEAQETEQAPKKGLLDALKVPLASIIPKRFKSATGEGEDDLELGKRPKNRAGLAS METLDDSLKDTETKDTVDNKPGANGTDSEALVKPEEKKDTEKDEADEPAPSKYP CLERIRNYRCTVDDIAIITGIVVFLLLLALIIAFTFIGKSDPVTAPIRDGKYIETVTSCG KVEGVLEDGAFAFRGIPYAVPPVGPLRWKAAQPIDNINYCWNETLKAHNSTPVC WQFYADGKVDGAEDCLTLDVITPNVRYDNPLPVVVLIGAESFTGDSPGKLRPSTR YARARDVIFVRPNFRLNVFGFLALEQLTKSSHPPTSGNYGLSDIIAALKWVQLNIA HFGGNPESVTLFGHRAGGTMVAALASSNKTSKLFARSWISSAASIYPGNPLPESEK

	ANTAYLGRIKCENAECLRDKEDEDVLDAVPDTWRRIFPDLPSVEENATANHEWL VLDGHILQQHPADVWSTETGKLKYVIGTTAHESHSEKLHLKYTEWTPELVTKHV NESKIGELGLTEEALKRYNATYQGLVAMISDIRTVCPLMTISQKLLTSQFYVVTQT GGELGIANVDSDIQAILGRYEPKTPEQRRYVSAIQQLFYHYVSHGEIKSELRRKLL DIGQDALPTYNSENCDFWIKNDIVPRYARLD
Anopheles gambiae	MGETDEKETAPAATTNDTTKPETPQPTADGPEVEAKETEKLLPEPEPKRSPVAQA GNPQSKAASSENVANGEEIINIPEEALADGADKDGKKDAAATEQEPTETGKPDKP NKVQAEEREVKPKKVPAGAFKLPGFFNKNKDKPKEADGADNELLEKHGGENGG EAKEKAAAEEGKPGEEKPKRTGGFFANLKLRNPFAKKPAAADATGETAADKEAG EEEEKDEVTAEATDKPADAEAAGDATEEKKPAAEGDEPEVQAPKKGLLEALRVP LASIIPKRFKPVPGGEPDDDIELGKTPKNRAGLASMETLDDSLKDADTKDTVDAK GANGTDGEALVKPDDKEAKDKAAEAEEDAQERSLLQRVQSYRCSVDDIAIIAGV VIFLILVALIIAFTFVGKSEPITAPVRDGKYIETVTNCGRVEGILEDGSFAFRGIPYAV PPVGPNRWKAAQPIESIENCWNGTLKAHNSTPVCWQIYADGKVDGAEDCLTLDV ITPHVRYDNPLPVVVLVGAESFTGDSPGKLRPSTRYARARDVIFVRPNFRLNVFGF LALEQLTKSTHPPTSGNYGLTDLIVALKWIQLNIPHFGGDPKSVTLFGHRAGGTIV TALASSNKTSKLFARTWISSGAAIFPGNPLAESEKENALYMSKIRCEDSNCLMEKE DEEILDAVPDVWRRTFPDLPAADENATARHEWLVLDGNIMQQHPADVWSADVS NSVRYVIGSTTHEAHNTKLHLKYTEWTPELVTRHVNESVVGRLGLTEEALRRYN ATYQGLVAMVSDIRTICPLLTITQKLQSSQFYVVSQTGGELGIADVDSDVQAILGR YEPKTPEQRRYVSAIQQLFYHYVSHGEIKHELRKKLLDIGQDALPTYNTDNCLFWI KNDVVPRYARLD
Culex quinquefasciatus	MGETDDKDTAPPTATEKGTVETPAAITEVQEPEPDAKETEKLLANSEEKTDDKET EKPTVEEKKKSSSENVNGTTNGEEIINIPDETSTETAKADDEKKPTDKPNKVQAEE REVKPKKVPAGEADGADNELLEKQDTEKEPKAVEAAAEEKPKRAGFFANLRLRN PFAKKQTDTAAADDTKTKPEEEENDEVTAEATDKKPDEETPATEATTKEEEAGV AAPKKGLLDALKVPLASIIPKRFKSGGGEGGEDDLELGKRPKNRAGLASMETLDD SLKDTETKDTVDKPAAAAANGTDSEALVKPDEKKDAEKGEDGEEPTPSKYPCWE RFRTYKCSVDDIAIIAGIVIFLLLLALIIAFTFIGRGEPATAPVRDGKYIETVTSCGKV EGILEDGAFAFRGIPYAVPPVGPLRWKAAQPIENINYCWNDTLKAHNSTPVCWQF YADGKVDGSEDCLTLDVITPHVRYDNPLPVVVLIGAESFTGDSPGKLRPSTRYAR ARDVIFVRPNFRLNVFGFLALDQLTKSVHPPTSGNYGLSDIIAALKWIQLNIAHFG GDPKSVTLFGHRAGGSIVAALASSNKTSKLFARSWISSASSIFPGNPLADSEAENSA YLGRIKCEDAECLREKEDEDVLDAVPDTWRRIFPDLPTSEENATANHEWLVLDGH ILQQHPADVWNTETGKIKYVVGSTIHESHSSKLYLKHTEWTPELVTKHINESKIGE LGLTEEAIKRYNATYKGLVAMISDIRTVCPLYTLSQKLLTSQFYVVSQTGGELNIA DVDSDIQAILGRYEPKTPEQRRYVSAIQQLFYHYVSHGEIKSELRRKLLDIGQDAL PTYNSDNCDFWIKNDVVPRYARLD
Tribolium castaneum	MDCHRFTANSQESEKFTERQPASGFFGSGQLLLPPAVRTKSPCLGRIAGQRGQLH AFLLCEELTVQASAGGTIEINGESFNSLHPQRVTPHEGCIMSDKIEGNVEKSDDKK DIEMEEREKMLNAENKALEKVEESADKPKTENKEGMEVKPKKIPIGGIQMPGFFT RSKSREKCKDEEQEVEGTELIETRAAGEAAPSQTRIKLPNPFRKSKGGEEDEKSAE QKEKKKLLDTIRLPLVSVFPRKKKEEQKLESQTAQAGLASMETLDDKSTNDELKN VNLDEKKDLEAQEVAEEATWRQKLKTYRVAIGALLVFLLLTLIIVIVAIPGKDVPR VAPIKDGGYVEAVTGCGKVEGLVEDSAVAFRGIPYARPPVGNLRFKYAQPLNNID YCWNGTFLAHNSTETCLQIHSNGSISGTEDCLTLDVVTPYVRYHDPLPVIVLIGAE SFIGGSPGKTRPSARYARSKDVVFVRPNFRLGALGFLALDILSNNDYPPTSGNYGL SDIIEALKWVQLNIRHFGGDNSSVTLLGHRAGATLVTALATLPDAKKYFSKAWAS SGGALYPNKTRREAEIDNRSFLQYVDCQNIDCLLKLDAEKIVRSVQDTWRKPPPD LPTPEEDPSKRHYWLTLDGHILKEHPATTWTQNKLKLVIGTTAHAGASETLLLKH KEWSEFLVRKHIHESFLTSRNLSESALEMYPRSYKGLSAMISDIRTVCPLMAITSQ MHEVPFYVVTQTRGEQDVADVDSDIDAILGRYEPKTPEQKRYFSAMQGLFYHYV WSDRIEHKDVTKKILLVGQDVLPNATYSHCAFWINKGVVLPFAALD

Nasonia vitripennis	MSQTDQTKQEQKPQEQEQQDKSLDKKEIAEEEREKMLNAENTKHTGAAPQAELE AEEQKPKRKIPIGGIKMPGFCRSKSKEPCKEEDTKPIETGDAETKESGDAAANDKA AAAGGGTPTKEAKEGRKGILDAIRLPLSSVFSRKKKESDAELGPTGAGLASSETLD DAAQEKNNAAAAVGEDGMETVRLDAEAADGPPTKPHPFILCLSAARRNLIVTGV TLLFLLTLIIIICVACVGPRRVVPQPIKDGKVQTQTSCGLVEGLVEEDAYVFRGIPY AVPPVNERRWQPSEQLRKIEHCWNGTYHAHNGSQVCLQRNPLNPSQSFNGSEDC LYLDVYTPQVMYERPLPVVVLIGADTLSGPSPGIMIPSGKLARVREVVYVRPNFRL DVFGFLAVEPLSRASHPMTSGNYGLSDLITALQWVQLNIEHFGGDKNLVTIWGHR AGGTLVTSLLGARKANGLFARAWVSSSSGMYPFEDRHDSEKKNEVFLKNINCND AACLKSKNADDIMKAVPSSWYTPNDAGLPDVKETDIKKHSWLVKDDIIVQQNPH TIFAQNKLPVKVMMGTTAQSGHLPSIFSYNKTEDPQQIRKVISDSVIGSNKELVDK ALAYYNTTVKGLHTMISDIRVVCPLQILANSNADVYFYVSTFPRNVVADVDSDAA AILGFFRAKTPEQKRHLQAIQQLFNHFVWHGEVKEEVRGSNRFFIIDQDVLPATNY DHCDQWISANLVPQNGRVD
Pediculus humanus	MERQDEIEQIRNDVEIIKTGEVEKMKKQLEASQKELGSSVRVPIGGIKIPGFLRSRK TKETDENSTEPDIEPPFVAQNVRKEDDKEPETQRQLIAKRFRKIITRMPLTKSGEGG GGGDDDDDNIENRDDAKDEKNLNTKKQKSFFNSIRIPLISDRPKKLNREFKNECSD IYLMFKKPLFMCLHLFCTYEAPNVGKTTKDMEMGTAGLASMETLEDSKESNGLV EDGMETVKLDVDDSKDVESQELASKRKQGVFGTWSVRDWRNWPEKMRRLTKH EKAIIAAIILFLILLILIIIILLAGSKQATSPPIKDGRYIVTSSSCGPIEGIFDEGIFKFHGIP YAVPPLGSKRWKYAEPLDQIDKCWNGTLKAYNRTSTCWQIYMEPNRKWRIDGA ENCLTLDIQTPFVRYDTPVPVIVLIDSGTLVGGSPGILKPTAKIIRSKEVVFVRPNFR LGIFGFLAADVLSKSTYPPVSGNYGLSDIITALKWIKTNIVHFGGDPNMITVLGHR AGATLTTALVSSHAAKNLFNRAWASSGSVIFPDKVLQLSEQANQRYIDRICAQKP DKLYCLKSYDAEELVEELEDSWRPTISDLPANYEANTTKHEWLVLDGNILKEGL WHDNSQQHTTLVLGTTAHSEASPQLWEKLQKIPGPKEEAIIEHVKGSYIGTMGLT NEALKLYNNSWEDLTAMISDIRTICPTLNFVRSQGAQQNVSFYVAKLERNHTDFG QIADIGVDMSSILGTYEPRTTAQRNFVSTIQDLFYNYVNHGEWTVKNREGVLIIGE KGEWFDDYPNCNFWISKNFVPKFARRD
Rhodnius prolixus	MMGSGPSKLVEFFKMGTLRWYIVCDRGLYFLFVPNIQNSRLYKYADEPYEGKTK TMTSKTLRRAIGFVKMSEFTGEGEKGSDVNLENKSEDKKDIEDEEKEKMIEEKEIR VGEVEKRKKALEEAAALGQVDKRRIPTGGIKIPGFLRSSRSRDKNKEGELEGEEEN DLIEPSCTAVRLGDEDKQNKTDLERKLPLIAKLNNINVPFLKRKKQEEGEEPSANG DAGNGEQDEPQRKPKLINAIRLPLASLVPKKHKLAKDGDHGVSSGPQAGLASME TLDDSNGSFKHDDGMENVRLDNVEFDPEKAALEDEVESKWATKEWNRLVKEHK IMIGSVILFLLLLILLLICLASGMQTPPPSAIINGRFITALTSCGLVEGVLEDGAHVFR GIPYALPPTGNLRFAPPKTFTLNDCWNGTLKVHNHTPPCWQMFVNGSLDGDEDC LTLDVYSPQVVVLIGADSLLGGWPGKFYPTAQIAVTNEVVFVHPHFRMGPLGFLA ARPLSDSVYPRTSGNYGLADVVQALLWVQLNIVHFGGDPKSVTVLGYRAGATLV TALTAMYKPERLFSRVWISSGSTIFPNISLVESESRTEGYLEAVSCNGNTTVATSCL RELDVEDLLDNIPDGWRGQEVGNLPKGQVKGHQWLVQDGVHIRQHPLDVWSSG NLSIQVVIGATAHSEINYNWKEMRSAWTPDRIKIEVNRSQIGAEGLTEKALGLYG NNLPGLAAMTSDIATICPLIKQHKSITNSKFYLVTQPADSFGLSYAGFDIDAILGTR FNATLVSAAQKYASLIRQMFFKYVHFGEQIPFILTSIGQNATQLNARDLEIKCNLW KEAGFHKFARIDFVQKNFKCMVWIPLSKL

Figure 11. Neurotactin Homologue Alignment

```
CLUSTAL 2.1 Multiple Sequence Alignments


Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: Dmel         846 aa
Sequence 2: Dana         844 aa
Sequence 3: Dgri         851 aa
Sequence 4: Dpso         845 aa
Sequence 5: Dvir         842 aa
Sequence 6: Dwil         859 aa
Sequence 7: Ccap         876 aa
Sequence 8: Mdom         843 aa
Sequence 9: Gmor         848 aa
Sequence 10: Aaeg         860 aa
Sequence 11: Agam         892 aa
Sequence 12: Cqui         857 aa
Sequence 13: Tcas         821 aa
Sequence 14: Nvit         736 aa
Sequence 15: Phum         808 aa
Sequence 16: Rpro         858 aa
```

---

clustalw.aln

```
CLUSTAL 2.1 multiple sequence alignment


Aaeg            -----------MGETDDKDTAPPTAPTEKPTTETPAAPEPHHEPEADPKETEKLLATREE
Cqui            -----------MGETDDKDTAPPTATEKG-TVETPAAITEVQEPEPDAKETEKLLANSEE
Agam            -----------MGETDEKETAPAATTNDTTKPETPQPTADG--PEVEAKETEKLLPEPEP
Mdom            -----------MGEQEEKEIPQSENSTQP-----EVAEAEPKETDKMLEKKDDKPSAEDT
Gmor            -----------MGELEEKETPATESTTQH-----ESAEE-PKETDKMLEKKEESIEPHSE
Ccap            -----------MGELEEKETGPSETTPQQ-----EVVDE-PKETDKMLDKKEDTEAVEQQ
Dmel            -----------MGELEEKETPPTETTAAQ-----QEALEEPKETDKMLDKKED--AKEKT
Dana            -----------MGELEEKETPPTETTTSQ-----QEALDEPKETDKMLDKKEDD-AKEKT
Dpso            -----------MGELEEQETPPTETTAAQ-----QETLEEPKETDKMLDSKKT----EKT
Dwil            -----------MGELEEKETPPTETTTNASSGHQQETLDEPKETDKMLDKKDETTSQEKT
Dgri            -----------MGELEEKETPQTETTTVQEST-----LDEPKETDKMLDAKK-----EKT
Dvir            -----------MGEIVEKETPQTEATAVQEP------LEEPKETDKMLDTKKDD-AKEKT
Phum            -----------MERQDEIEQIRN-----------------DVEIIKTGEVEKMKKQLEAS
Tcas            MDCHRFTANSQESEKFTERQPASGFFGSGQLLLPPAVRTKSPCLGRIAGQRGQLHAFLLC
Nvit            ---------------------------------------------MSQTDQTKQEQKPQEQ
Rpro            ---MMGSGPSKLVEFFKMGTLRWYIVCDRGLYFLFVPNIQNSRLYKYADEPYEGKTKTMT


Aaeg            KTDDKETEKPADEKKRSSENVNG-----TEEIINIPDETSTETVKVEDEK----------
Cqui            KTDDKETEKPTVEEKKKSSSENVNGTTNGEEIINIPDETSTETAKADDEK----------
Agam            KRSPVAQAGNPQSKAASSENVAN-----GEEIINIPEEALADGADKDGKKDAAAT-----
Mdom            KSKNGSKPTTPTATEKTEKKTPAGSPTSEKKKEEIIDIPEEENNDE--------------
Gmor            KSADNKKKSNSKPTTPVSERKTSEEMRSNTSEEIIDVPVESDAKTT--------------
Ccap            TENAGVKDKNPSPNTEKKRAASSKPATPTSLGGKKSSIQSVEKLATNGVDDAQSVKANGG
Dmel            PSPQTSKPASPNAGKKSSPVAEKKIDDAELAKSKSGNGEEIIDIPAENGTKP--------
Dana            PSPQTSKPASPNAGKKSSPLAEKKAVEGDTQK--PGNGEEIIDIPAENGTKS--------
Dpso            PSPQSSKPATPNAGKKSPVGEKTNEEKGHEIIDIPEEKEKPQPTEGSG------------
Dwil            ATPHNSKPSTPDAKKKLENDEKK-----------SNGGEEIIDIPNENGTKP---------
Dgri            PSPQTSKPGTPNTGKKSPEVPLEVLSKNVVEKP-LNNGEEIIDIPTENGNKPG-------
Dvir            ASPQTSKPASPNAGQKSPTLPQADPAEKPAGG----NGEEIIDIPTENGNKP--------
Phum            QKELGSSVRVPIGGIKIPGFLRSRKTKETDENSTEPDIEPP--------------------
Tcas            EELTVQASAGGTIEINGESFNSLHPQRVTPHEGCIMSDKIEGNVEKSDDKKDIEMEEREK
Nvit            EQQDKSLDKKEIAEEEREKMLNAENTKHTG------------------------------
Rpro            SKTLRRAIGFVKMSEFTGEGEKGSDVNLENKSEDKKDIEDEEKEKMIEEKEIRVG-----


Aaeg            -------------AKTDKPNKVQAEEREVKPKKVPAGAFKLPGFFNKGK--AKEADGADN
Cqui            -------------KPTDKPNKVQAEEREVKPKKVPAG----------------EADGADN
Agam            ------EQEPTETGKPDKPNKVQAEEREVKPKKVPAGAFKLPGFFNKNKDKPKEADGADN
```

```
Mdom            ------ATKTNGAGGEGDERKMSGEEREVKPKKIPIGGLKLPGFFVKNKP-KGEGDGAEG
Gmor            ------IDEEN--HDNGDDKKISTEDREVKPKKIPIGGLKLPGFFMKSKP-RVDGDGADG
Ccap            SGEEIIDIPESTKTEEGDEKKISTEDREVKPKKIPIGGLKLPGFFMKNKP-KADGDGADG
Dmel            --------------DSADDKKISKEEREVKPKKIPIGGLKLPGFFMKNKP-KADGDGAEG
Dana            --------------SEGDDKKISKEEREVKPKKIPIGGLKLPGFFMKNKP-KAEGDGAEG
Dpso            ---------------AGEEKKISKEEREVKPKKIPIGGLKLPGFFMKNKP-KAEGDGAEG
Dwil            --------------SENEDKKISKEEREVKPKKIPIGGLKLPGFFMKNKP-KSEGDGAEG
Dgri            -----------AAIDVTDEKKISKEEREVKPKKIPIGGLKLPGFFMKNKP-KGEGDGAEG
Dvir            -------------NGAADEKKVSKEEREVKPKKIPIGGLKLPGYFMKNKP-KAEGDGAEG
Phum            -------------FVAQNVRKEDDKEPETQRQLIAKRFRKIITRMPLTKS----GEGGGG
Tcas            MLNAENKALEKVEESADKPKTENKEGMEVKPKKIPIGGIQMPGFFTRSKS---------R
Nvit            ----------------AAPQAELEAEEQKPKRKIPIGGIKMPGFCRSKSK------EPCK
Rpro            -----------EVEKRKKALEEAAALGQVDKRRIPTGGIKIPGFLRSSRSRDKNKEGELE
                                         :   : :.

Aaeg            ELLEKQETDKDAKAAEAAEE-KPKRP-----------GFFANLRLRNPFAKKQAEPVPTE
Cqui            ELLEKQDTEKEPKAVEAAAEEKPKRA-----------GFFANLRLRNPFAKKQTDTAAAD
Agam            ELLEKHGGENGGEAKEKAAAEEGKPGEEKPKR---TGGFFANLKLRNPFAKKPAAADATG
Mdom            ELLEKEPKE---EEAKTAEEKTSKKEEKPG--------KNFGQRLRSFFVRKPAAEKQAK
Gmor            ELLEKEHKE---ETTS--EVKLAKKDEKPG--------KNFGKRLRDFFVRKPAAEKQAK
Ccap            ELLERENKDEMPAETVNGESKSSKKDEKPR--------SSFGERLRNFFVRKPTADKQQN
Dmel            ELLEKEKEED--KDKEANGDAATGSGKDEQ-----KSRPGLGERLRSFFARKPSAEKEKK
Dana            ELLEKEKEED--KDKEANGDAATTSGKDEQ-----KARPGLGERLRSFFVRKPAAEKEKK
Dpso            ELLEKDAEDKDKEATAVNGDAAAAAAAAEP-----KPRPGLGQRLRSFFVRKPGAEKEKK
Dwil            ELLEKEKDDDENSKEAKVENGEAKGKDDQQQQQQQKSRPGLGERLRNLFVRKPATDKEKK
Dgri            ELLEKDNEQEQEEKEANGGDATTTPNKDEQ-----KPRTNLGQRLRSFFVRKPAADKEQQ
Dvir            ELLEKDKELAEEK------EANGDASKEEQ-----KPRTGLGQRLRSLFVRKPAADKEQQ
Phum            GDDDDDNIENRDDAKDEKNLNTKKQKSFFN-----------SIRIPLISDRPKKLNREFK
Tcas            EKCKDEEQEVEGTELIETRAAGEAAPSQTR------------IKLPNPFRKSKGGEEDEK
Nvit            EEDTKPIETGDAETKESGDAAANDKAAAAG-------------------------GGTPTK
Rpro            GEEENDLIEPSCTAVRLGDEDKQNKTDLER----KLPLIAKLNNINVPFLKRKKQEEGEE


Aaeg            ETKAKPEE---EENDEVTAEATDKKPEDES--AVEKKEE-----AQETEQAPKKGLLDAL
Cqui            DTKTKPEE---EENDEVTAEATDKKPDEETPATEATTKE-----EEAGVAAPKKGLLDAL
Agam            ETAADKEAGEEEEKDEVTAEATDKPADAEAAGDATEEKKPAAEGDEPEVQAPKKGLLEAL
Mdom            QAASN---------------EADNKSGKFFPSLFSPTFN------------CKRGLLNAI
Gmor            QQAAS---------------TGDADTKSEATAEAGPASEGNMR-TNSDGPPPKRGLLNAI
Ccap            KQTANG--------------DADAKSEATAEAAAADGETNNTT-NASDAPPQKRGLLNAI
Dmel            QLVNG-----------------DADAKSEATAEATPAEDA-----SDAPP--KRGLLNAI
Dana            QLVNG-----------------DADAKSEATAEAAPAEDA-----AEAPP--KRGLLNAI
Dpso            ALVNG-----------------DADAKSEATAEAAAAEDA-----AAVDPPPKRGLLNAI
Dwil            QQLINGKE------------TADNDNKSEATAEANGQEES-----SVDPPP-KRGLLNAI
Dgri            KKQLN---------------NGDADAKSEATAEAAAVED--------ADPPPKRGLLNAI
Dvir            KKPLT---------------NGDADAKSEATAEAAPGED--------ADPPPKRGLLNAI
Phum            NECSD-------------------------------------------IYLMFKKPLFMCL
Tcas            SAEQK-------------------------------------------EKKKLLDTI
Nvit            EAKEG-------------------------------------------RKGILDAI
Rpro            PSANGDAGN----------------------------------GEQDEPQRKPKLINAI
                                                             :  ::  :

Aaeg            KVPLASIIPKRFKSATGEG-EDDLELGKRPKNRAGLASMETLDDSLKDTETKDTVDN--K
Cqui            KVPLASIIPKRFKSGGGEGGEDDLELGKRPKNRAGLASMETLDDSLKDTETKDTVDKPAA
Agam            RVPLASIIPKRFKPVPGGEPDDDIELGKTPKNRAGLASMETLDDSLKDADTKDTVDAK--
Mdom            KLPIANMVPK-------KKSDDDVELGMG---KAGLASMETLDDSLKDQDVVDKAAVTKN
Gmor            KLPIANMIPK-------KSPDDDVELGMG---KAGLASMETLDDSLKDQDCVDKASPKNN
Ccap            KLPIANMIPR-------KKTDDDVELGMG---KAGLASMETLDDSLKDQDCVDKAPARNN
Dmel            KLPIANMIPK-------KKSNDDVELGLG---KAGLASMETLDDSLKDQDTVDRAPVKTN
Dana            KLPIANMIPK-------KKNNDDVELGLG---KAGLASMETLDDSLKDQDTVDRAPVKTN
Dpso            KLPIANMIPKQ------RKNGDDVEMGLGG--KAGLASMETLDDSLKDQDTVDRAPVKNA
Dwil            KLPIANMIPK-------KKSPDDVELGLGGGNKAGLASMETLDDSLKDQDTVDRAPVKSN
Dgri            KLPIANMMPR-------KKAGDDVELGLG---KAGLASMETLDDSLKDQDTVDRAPLKTN
Dvir            KLPIANIMPK-------KRGGDDVEMGLG---KAGLASMETLDDSLKDQDTVDRAPLKSN
Phum            HLFCTYEAPN------VGKTTKDMEMGTAG-----LASMETLEDSKESNGLVEDGMETVK
Tcas            RLPLVSVFPR---------KKKEEQKLESQTAQAGLASMETLDDKSTNDELKN-------
Nvit            RLPLSSVFSR-------KKKESDAELGPTG---AGLASSETLDDAAQEKNNAAAAVGEDG
Rpro            RLPLASLVPKK------HKLAKDGDHGVSSGPQAGLASMETLDDSNGSFKHDDGMENVRL
                ::      ..           .: :          *** ***:*   .

Aaeg            PGANGTDSEALVKPEEKKDTEKDE-ADEPAPSKYPCLERIRNYRCTVDDIAIITGIVVFL
Cqui            AAANGTDSEALVKPDEKKDAEKGEDGEEPTPSKYPCWERFRTYKCSVDDIAIIAGIVIFL
```

```
Agam            -GANGTDGEALVKPDDKEAKDKAAEAEEDAQERS-LLQRVQSYRCSVDDIAIIAGVVIFL
Mdom            GNDDVGKLKKSPSGEIKQASSDEKSPDGDEEQPISFCQRLRSYKCSVDDALIALGILLFV
Gmor            GVDEISKLKKTPSGEIKQAS-EEKTFSDAEDVQISFCQRLRSYKCSVDDALIALGILLFI
Ccap            GTEDPAKLKKATSSEIKQESPEDKATEEPE-APMSFADRLRSYRCSVDDGLIVLGILLFL
Dmel            ---GTEELKGELKDEKLAAEEKLAAEEEEQNRPVSLLTRLRGYKCSVDDALIVFGILLFV
Dana            ---GTDEIKSDLKDEKLAAEQKLAAEEEERERPVPLLTRLRGYKCSVDDALIVFGILLFV
Dpso            N--GTDDIKSELKDEKLAAEQKLAAEEDERDRPISLLQRLRGYKCSVDDALIAFGILLFV
Dwil            G--TDDLKSIELKDDKLPANEKLAAEEEEQERPASLWTRLRGYKCSVDDALIVFGILLFV
Dgri            ---GADELKSDLKDDKLAAEQKLAAEEEERDRPISLWCRLLGYKCSVDDALIVFGILLFM
Dvir            ---GADELKSELKSDKLAAQQKLAAEEEERERPVSVWCRLRGYKCSVDDALIVFGILLFM
Phum            --LDVDDSKDVESQELASKRKQGVFGTWSVRDWRNWPEKMR--RLTKHEKAIIAAIILFL
Tcas            ----------------VNLDEKKDLEAQEVAEEATWRQKLKTYRVAIG------ALLVFL
Nvit            ------------------METVRLDAEAADGPPTKPHPFILCLSAARRNLIVTGVTLLFL
Rpro            DN-----------------VEFDPEKAALEDEVESKWATKEWNRLVKEHKIMIGSVILFL
                                                                     ::*:

Aaeg            LLLALIIAFTFIGKSDPVTAPIRDGKYIETVTSCGKVEGVLEDGAFAFRGIPYAVPPVGP
Cqui            LLLALIIAFTFIGRGEPATAPVRDGKYIETVTSCGKVEGILEDGAFAFRGIPYAVPPVGP
Agam            ILVALIIAFTFVGKSEPITAPVRDGKYIETVTNCGRVEGILEDGSFAFRGIPYAVPPVGP
Mdom            LLIAVIGYVLSHEG--QTSPPIREGRFMDAVTGCGMVEGLKEDGAFAFRGIPYAVPPVGD
Gmor            FLIAVIGYVLSYET--LTSPPIREGRFMDAVTGCGMVEGLKEDGAFAFRGIPYALPPTGD
Ccap            ALIAVIGYVLSLET--LTSPPVREGRFIDAVTGCGMVEGLQEDGAFAFRGIPYALPPLGD
Dmel            LLLGVIGYVLTHET--LTSPPLREGRYIMAVTGCGPVEGVKEDGAFAFRGIPYAKPPVDR
Dana            LLLAVIGYVLTHET--LTSPPLREGRYIVAVTSCGPVEGVKEDGAFAFRGIPYAKPPVDG
Dpso            LLLAVIGYVLTHET--LTSPPLREGRYILAVTGCGPVEGVKEDGAFAFRGIPYAKAPVDE
Dwil            LLMAVIGYVLTHET--LTSPPLREGKFILAVTGCGPVEGVKEDGAFAFRGIPYAKAPIEE
Dgri            LLLAVIGYVLTHET--LTSPPLREGRYISAVTSCGMVEGVKEDGAFAFRGIPFAKAPIDD
Dvir            LLLAVIGYVLTHET--LTSPPLREGRYISAVTSCGMVEGVKEDGAFAFRGIPYAKAPLEN
Phum            ILLILIIIILLAGSK-QATSPPIKDGRYIVTSSSCGPIEGIFDEGIFKFHGIPYAVPPLGS
Tcas            LLTLIIVIVAIPGKDVPRVAPIKDGGYVEAVTGCGKVEGLVEDSAVAFRGIPYARPPVGN
Nvit            LTLIIIICVACVGPRRVVPQPIKDG-KVQTQTSCGLVEGLVEEDAYVFRGIPYAVPPVNE
Rpro            LLLILLLICLASGMQTPPPSAIINGRFITALTSCGLVEGVLEDGAHVFRGIPYALPPTGN
                    ::              .: :*  : : :.** :**: ::.   *:***:* .*

Aaeg            LRWKAAQPIDNINYCWNETLKAHNSTPVCWQFYADG----KVDGAEDCLTLDVITPNVRY
Cqui            LRWKAAQPIENINYCWNDTLKAHNSTPVCWQFYADG----KVDGSEDCLTLDVITPHVRY
Agam            NRWKAAQPIESIENCWNGTLKAHNSTPVCWQIYADG----KVDGAEDCLTLDVITPHVRY
Mdom            RRWRPAQTIESINDCWNGTLKAHNTSEVCTQILGNG----TVVGDEDCLYLDVVTPHVRY
Gmor            RRWRSAQTIEGIEDCWNGTLQAHNSSAMCTQMLGNG----TVIGDEDCLYLDVITPHIRY
Ccap            QRWRPAQPIGGIDDCWNGTLKAHNVSNYCTQRLGNG----TIVGDEDCLYLDIVTPHVRY
Dmel            LRWKPAELIDDINMCWNDTLQTHNSSVVCTQRLGNG----TTVGDEDCLYLDVVTPHVRY
Dana            LRWRPAELVDSIDGCWNDTLQTHNSSVVCTQRLGNG----TTVGEEDCLYLDVVTPHVRY
Dpso            LRWRPAQLIDSIDGCWNDTLQTHNSSVVCTQRLGNG----TTVGDEDCLFLDVVTPHVRY
Dwil            RRWRPAQLIDSIDGCWNDTLQTHNSSVVCTQRLGNG----TTVGDEDCLYLDVVTPHVRY
Dgri            LRWRPAQLIEKIDDCWNGTLQTHNESTVCTQRLGNG----TTIGDEDCLYLDVVTPHVRY
Dvir            LRWRPAQLISKIDDCWNGTLQTHNGSAVCMQRLGNG----TTIGDEDCLYLDVVTPHVRY
Phum            KRWKYAEPLDQIDKCWNGTLKAYNRTSTCWQIYMEPNRKWRIDGAENCLTLDIQTPFVRY
Tcas            LRFKYAQPLNNIDYCWNGTFLAHNSTETCLQIHSNG----SISGTEDCLTLDVVTPYVRY
Nvit            RRWQPSEQLRKIEHCWNGTYHAHNGSQVCLQRNPLNPSQ-SFNGSEDCLYLDVYTPQVMY
Rpro            LRFAPPKTFT-LNDCWNGTLKVHNHTPPCWQMFVNG----SLDGDEDCLTLDVYSP----
                 *:  .: .  :: *** *  .:* :  * *            * *:** **: :*

Aaeg            DNPLPVVVLIGAESFTGDSPGKLRPSTRYARARDVIFVRPNFRLNVFGFLALEQLTKSSH
Cqui            DNPLPVVVLIGAESFTGDSPGKLRPSTRYARARDVIFVRPNFRLNVFGFLALDQLTKSVH
Agam            DNPLPVVVLVGAESFTGDSPGKLRPSTRYARARDVIFVRPNFRLNVFGFLALEQLTKSTH
Mdom            ENPLPVVVMIGAETLMGPSPGIVRPTARFSRSHDVIFVRPNFRLGPFGFLALEMLTKDTY
Gmor            DSPLPVIVMISAESLTGPSPGVLRPSARYSRSHDVIYVRPNFRLGAFGFLAVEVLTKDSY
Ccap            NSPVPVIVLIGAETLTGPSPGVLRPSARYSRSHDVIFVRPNFRLGAFGFLALEALTKDAY
Dmel            NNPLPVVVLIGAESLAGPSPGILRPSARYSRSHDVIFVRPNFRLGVFGFLALDALTKEAH
Dana            TNPLPVVVLIGAESLAGPSPGILRPSARYSRSHDVIFVRPNFRLGVFGFLALDALTKETR
Dpso            TNPLPVVVLIGAETLAGPSPGILRPSARYSRSHDVIFVRPNFRLGIFGFLALESLTKEAH
Dwil            TNPLPVVVLIGAESLAGPSPGILRPSARYSRSHDVIFVRPNFRLGIFGFLALDALTKESH
Dgri            TNPLPVVVLIGAESLTGPSPGILRPSARYSRSHDVIFVRPNFRLGVFGFLALDALTKESH
Dvir            TNPLPVVVLIGAESLAGPSPGILRPSARYSRSHDVIFVRPNFRLGVFGFLALDALTKESH
Phum            DTPVPVIVLIDSGTLVGGSPGILKPTAKIIRSKEVVFVRPNFRLGIFGFLAADVLSKSTY
Tcas            HDPLPVIVLIGAESFIGGSPGKTRPSARYARSKDVVFVRPNFRLGALGFLALDILSNNDY
Nvit            ERPLPVVVLIGADTLSGPSPGIMIPSGKLARVREVVYVRPNFRLDVFGFLAVEPLSRASH
Rpro            ----QVVVLIGADSLLGGWPGKFYPTAQIAVTNEVVFVHPHFRMGPLGFLAARPLSDSVY
                     *:*::.: :: *  **   *: :    .:*::*:*:**:. :****   *:

Aaeg            PPTSGNYGLSDIIAALKWVQLNIAHFGGNPESVTLFGHRAGGTMVAALASSNKTSKLFAR
```

```
Cqui            PPTSGNYGLSDIIAALKWIQLNIAHFGGDPKSVTLFGHRAGGSIVAALASSNKTSKLFAR
Agam            PPTSGNYGLTDLIVALKWIQLNIPHFGGDPKSVTLFGHRAGGTIVTALASSNKTSKLFAR
Mdom            PHTSGNYALSDIIAALNWIHLNIRHFGGDPKSVTLLGHRAGATLVSALITSKKAEGLFTR
Gmor            PHTSGNYALSDIIAALKWIKLNIVHFGGDPKSVTILGHRAGGTLVTALITSKKVEGLYTR
Ccap            PRTSGNYALSDIIAALKWIQLNIAHFGGDPKSVTLLGHRAGGTLVSALVTSNKVEGLYKN
Dmel            PPTSGNYALTDIIAVLNWIKLNIVHFGGDPQSVTLLGHRAGATLVTLLVNSQKVKGLYTR
Dana            PPTSGNYALTDIIAVLKWIKLNIANFGGDPKSVTLLGHRAGATLVTALVTSNKVKDLYTR
Dpso            PRTSGNYALTDIIAVLKWIQLNIVNFGGDPKSVTLLGHRAGATLVSALVSSPKLKDLYTR
Dwil            PRTSGNYALTDIIAVLKWVQLNIVNFGGDPKSVTLLGHRAGATLVTALVSSPKLQGLYTR
Dgri            PRTSGNYALTDIVAALKWIQRNIMHFGGDPKSVTLLGHRAGATLVSTLVSSAKLEEYYTR
Dvir            PRTSGNYALTDIIAVLKWIQLNIVHFGGDPKSVTLLGHRAGATLVSVLVSSAKLEGLYTR
Phum            PPVSGNYGLSDIITALKWIKTNIVHFGGDPNMITVLGHRAGATLTTALVSSHAAKNLFNR
Tcas            PPTSGNYGLSDIIEALKWVQLNIRHFGGDNSSVTLLGHRAGATLVTALATLPDAKKYFSK
Nvit            PMTSGNYGLSDLITALQWVQLNIEHFGGDKNLVTIWGHRAGGTLVTSLLGARKANGLFAR
Rpro            PRTSGNYGLADVVQALLWVQLNIVHFGGDPKSVTVLGYRAGATLVTALTAMYKPERLFSR
                * .****.*:*:: .* *:: ** :***: . :*: *:***.::.: *      .  : .

Aaeg            SWISSAASIYPGNPLPESEKANTAYLGRIKCENAE-----CLRDKEDEDVLDAVPDTWRR
Cqui            SWISSASSIFPGNPLADSEAENSAYLGRIKCEDAE-----CLREKEDEDVLDAVPDTWRR
Agam            TWISSGAAIFPGNPLAESEKENALYMSKIRCEDSN-----CLMEKEDEEILDAVPDVWRR
Mdom            AWISSASAILPGKPLSESEKKNEELAQALDCKDVA-----CMRNVEAEKLWDAIPDTWLH
Gmor            AWISSASAIYPGKSLVESEKRNQEFLQALECQDVK-----CLRAASAEKIWDATPDTWLH
Ccap            AWVSSSSAIFPGKPLEESEKRNAQFMTALDCSSIS-----CLRNAATEQIWAATPDTWLH
Dmel            AWASSGSAILPGKPLSESGKQNEQLMATLECADIQ-----CLREASSERLWAATPDTWLH
Dana            AWASSGSAILPGRPLSVSENLNQQLMGTLECADVE-----CLREASSERLWAATPDTWLH
Dpso            AWSSSGSAIFPGKPLAEAEKLNQQLMVTLDCSDVG-----CLRDASSEKLWAATPDTWLH
Dwil            AWASSASAIFPGKPLSESEKVNQQLMATLDCQDVE-----CLREASTEKLWAATPDTWLH
Dgri            AWASSGSAIFPGKPLHESERSNQQLMTTLECQDAV-----CLRETSSEKLWAATPDTWFH
Dvir            AWASSASAIYPGKPLQESERLNQQLMATLECQDAE-----CLRAASSEKLWAATPDTWFQ
Phum            AWASSGSVIFPDKVLQLSEQANQRYIDRICAQKPDKLY--CLKSYDAEELVEELEDSWRP
Tcas            AWASSGGALYPNKTRREAEIDNRSFLQYVDCQNID-----CLLKLDAEKIVRSVQDTWRK
Nvit            AWVSSSSGMYPFEDRHDSEKKNEVFLKNINCNDAA-----CLKSKNADDIMKAVPSSWYT
Rpro            VWISSGSTIFPNISLVESESRTEGYLEAVSCNGNTTVATSCLRELDVEDLLDNIPDGWRG
                 * **.. : *      :   .      : .         *:     : :     . *

Aaeg            IFPDLPSVEENATAN--HEWLVLDGHILQQHPADVWSTETG-----KLKYVIGTTAHESH
Cqui            IFPDLPTSEENATAN--HEWLVLDGHILQQHPADVWNTETG-----KIKYVVGSTIHESH
Agam            TFPDLPAADENATAR--HEWLVLDGNIMQQHPADVWSADVSN----SVRYVIGSTTHEAH
Mdom            FPADIPAAEENTTA--HHEWLVLDGDILKEHPAESWKKEHVG----KPLLVMGTTAHESH
Gmor            FPVDIPTAVENKTA--HHEWLVLDGDILQEHPAESWKKEHAG----DPVLVMGTTAHESH
Ccap            FPVDMPTTDENPST--RHEWLVLDGNILQRHPADAWKEEHTG----KPKLVMGTTAHEAH
Dmel            FPVDLPQPQEANASGSRHEWLVLDGDVVFEHPSDTWKREQAN---DKPVLVMGATAHEAH
Dana            FPVDIPQPQEFNAG-QRHEWLVLDGDVLYQHPSESWKQDKAT---DKPLLVMGATAHEAH
Dpso            FPQDLPQPQEINSG-SRHEWLVLDGDVLFQHPSESWKRDQAT---EQPLLVVGATAHEAH
Dwil            FPMDLPQRQELSSSGQRHEWLVLDGDILYQHPSESWKLEQAT---DKPLLVMGATAHEAH
Dgri            FPTTSPQ-ELELSNANRHEWLVLDGDILQQHPSESWKQVPTASQLQHPRYVMGATAHEAH
Dvir            FPATVPQ-AEELSNANRHEWLVLDGDILYQHPSESWKQVPTASQLHQPRLVMGATAHEAH
Phum            TISDLPANYEANTT--KHEWLVLDGNILKEGLWHDNSQQHTT-------LVLGTTAHSEA
Tcas            PPPDLPTPEEDPSK--RHYWLTLDGHILKEHPATTWTQNKLK-------LVIGTTAHAGA
Nvit            -PNDAGLPDVKETDIKKHSWLVKDDIIVQQNPHTIFAQNKLP-----VKVMMGTTAQSGH
Rpro            ----QEVGNLPKGQVKGHQWLVQDGVHIRQHPLDVWSSGNLS-----IQVVIGATAHSEI
                             * **. *.  : .                   ::*:* :

Aaeg            SEKLHLKYTEWT---PELVTKHVNESKIGELGLTEEALKRYNAT-YQGLVAMISDIRTVC
Cqui            SSKLYLKHTEWT---PELVTKHINESKIGELGLTEEAIKRYNAT-YKGLVAMISDIRTVC
Agam            NTKLHLKYTEWT---PELVTRHVNESVVGRLGLTEEALRRYNAT-YQGLVAMVSDIRTIC
Mdom            TESLHKRFGNWT---AEEIRSFLENSKIGALNLTDEAIQRYNATNYRALVAMITDIRTVC
Gmor            TEKLRRRFANWT---AEEIRNFIESSKIGTLNLTDEAIRRYNATNYPALVSMITDIRTIC
Ccap            TEQLRQRHANWT---SEEIRAFIATSKIGALNLTDEAIRLYNATNYRSLVAMITDIRTIC
Dmel            TEKLRELHANWT---REEVRAYLENSQIGALGLTDEVIEKYNASSYASLVSIISDIRSVC
Dana            SERLRELHANWT---REEVRAYLTNSQIGAMGLTEEVLERYNASSYASLVSIITDIRTVC
Dpso            TEKLRELHANWT---KEKVRDYLNNSQIGALGLTDEVIERYNATNYAALVSIITDIRTVC
Dwil            SLKLRELHANWT---KAEIQDYLNKSQLGAKSLTDEVMQLYNATNYASLLAIITDIRTVC
Dgri            TLKLRERHANWT---AEEVRSYLQASKLGELGLTDEIIQRYNATNYASLVAIITDIQTVC
Dvir            TLKLRELHANWT---AEEVRSYLQNSQLGARGLTDEIIQRYNATSYAALVAIITDIRTVC
Phum            SPQLWEKLQKIPGPKEEAIIEHVKGSYIGTMGLTNEALKLYNNS-WEDLTAMISDIRTIC
Tcas            SETLLLKHKEWS---EFLVRKHIHESFLTSRNLSESALEMYPRS-YKGLSAMISDIRTVC
Nvit            LPSIFSYNKTED---PQQIRKVISDSVIGSNKELVDKALAYYNTTVKGLHTMISDIRVVC
Rpro            NYNWKEMRSAWT---PDRIKIEVNRSQIGAEGLTEKALGLYGNN-LPGLAAMTSDIATIC
                             :   :  * :       .    *  .    * :: :**  :*
```

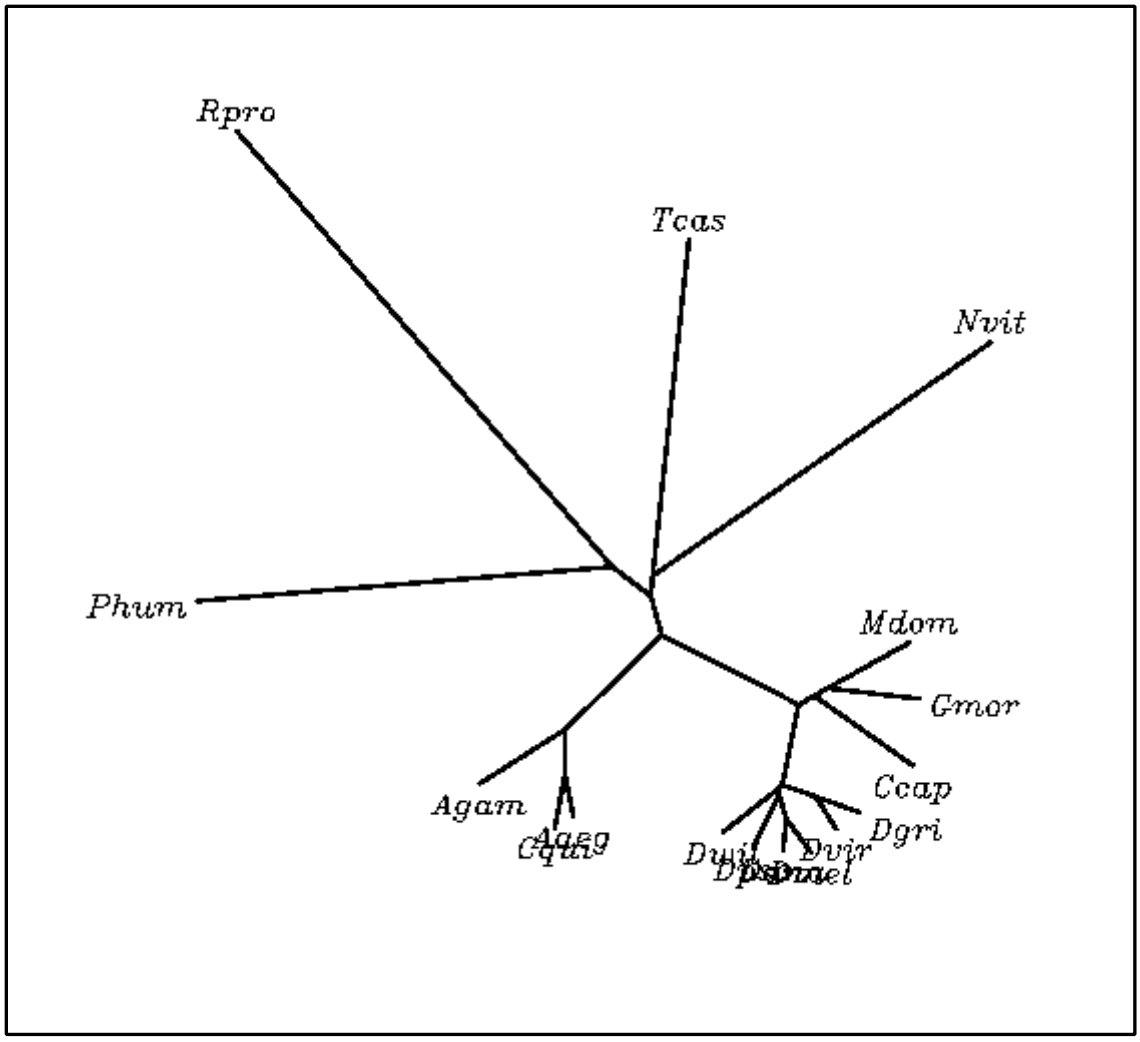
```
Aaeg            PLMTISQ---KLLTSQFYVVTQTG---GELGIANVDSDIQAILGRYEPKTPEQRRYVSAI
Cqui            PLYTLSQ---KLLTSQFYVVSQTG---GELNIADVDSDIQAILGRYEPKTPEQRRYVSAI
Agam            PLLTITQ---KLQSSQFYVVSQTG---GELGIADVDSDVQAILGRYEPKTPEQRRYVSAI
Mdom            PLLINAR---LQPSVPFYVVSQGE---GEEQLATVDADIQAILGRYEPHTVEQRRFVSSM
Gmor            PLLINAR---LQPSAPFYVVTQGA---GEEQLATVDADVQAIMGRYEPHTVEQRRFVSSM
Ccap            PLLTNAR---LQPTVPFYVVQQGE---GENQLATVDTDVLAILGRYEPHTAEQRRFMSSM
Dmel            PLLTNAR---QQPSVPFYVVTQGE---GPDQLATVDADVQAILGRYEPHTVEQRRFVSAM
Dana            PLLTNAR---LQLGVPFYVVTQGE---GPDQLATVDADVQAILGRYEPHTVEQRRFVSAM
Dpso            PLLTNAR---LQPSVPFYVVTQGE---GEDQLATVDADVQAILGRYEPHTVEQRRYVSAM
Dwil            PLLANAR---LQPSVPFYVVTQGE---GDDQLANVDADVQAILGRYEPHTVEQRRFVSAM
Dgri            PLLTNAR---LQPSVPFYVVTQGE---GDDQLATVDADVQAILGRYEPHTVEQRRFVSAM
Dvir            PLLTNAR---LQPTVPFYVVTQGE---GDDQLATVDADVQAILGRYEPHTVEQRRFVSAM
Phum            PTLNFVRSQGAQQNVSFYVAKLERNHTDFGQIADIGVDMSSILGTYEPRTTAQRNFVSTI
Tcas            PLMAITS---QMHEVPFYVVTQTR---GEQDVADVDSDIDAILGRYEPKTPEQKRYFSAM
Nvit            PLQILAN---SNADVYFYVSTFPR-----NVVADVDSDAAAILGFFRAKTPEQKRHLQAI
Rpro            PLIKQHKS---ITNSKFYLVTQPADS-FGLSYAGFDIDAILGTRFNATLVSAAQKYASLI
                *               **:            * .. *         . .   :.. . :

Aaeg            QQLFYHYVSHGEIKSELR--RKLLDIGQDALPTYNSENCDFWIKNDIVPRYARLD-----
Cqui            QQLFYHYVSHGEIKSELR--RKLLDIGQDALPTYNSDNCDFWIKNDVVPRYARLD-----
Agam            QQLFYHYVSHGEIKHELR--KKLLDIGQDALPTYNTDNCLFWIKNDVVPRYARLD-----
Mdom            QQLFYYYVAHGNVPSYDPR-RRVIDVGQDPLSQEDHANCNFWISNDIVPRYARVD-----
Gmor            QQLFYYYVSHGTVPQYDPR-RRVINVGQDPLSQSDYPNCNFWISNDIVPRYARID-----
Ccap            QQLFYYYISHGTVPQYDPS-RRVLSIAQDPLPQSDHPNCNFWISHDIVPRYARVD-----
Dmel            QQLFYYYVSHGTVQSFVQN-RRVINVGQDAQPEEDYLPCNYWISKDIVPRYARVD-----
Dana            QQLFYYYVSHGTVQTFEPQ-RRVINVGQDAQSQEDFEPCNYWISKDIVPRYARVD-----
Dpso            QQLFYYYVSHGTVQSFEPR-RRVINVGQDAQPQEDYEHCNYWISKDIVPRYARID-----
Dwil            QQLFYYYVSHGTVQSYDSR-RRVINVGQDAQPQEDHKNCNYWISKDIVPRYARID-----
Dgri            QQLFYYYVSHGTVQTFESR-RRVIIVGQDAQSQEDLEHCNYWISKDIVPRYARVD-----
Dvir            QQLFYYYVSHGTVQSFDTR-RRVIIVGQDAQAQEDHDHCNYWISKDIVPRYARVD-----
Phum            QDLFYNYVNHGEWTVKNR--EGVLIIGEKGEWFDDYPNCNFWISKNFVPKFARRD-----
Tcas            QGLFYHYVWSDRIEHKDVT-KKILLVGQDVLPNATYSHCAFWINKGVVLPFAALD-----
Nvit            QQLFNHFVWHGEVKEEVRGSNRFFIIDQDVLPATNYDHCDQWISANLVPQNGRVD-----
Rpro            RQMFFKYVHFGEQIPFILT--SIGQNATQLNARDLEIKCNLWKEAGFH-KFARIDFVQKN
                : :*  ::  .           .     .         *  * . ..    .  *

Aaeg            ------------
Cqui            ------------
Agam            ------------
Mdom            ------------
Gmor            ------------
Ccap            ------------
Dmel            ------------
Dana            ------------
Dpso            ------------
Dwil            ------------
Dgri            ------------
Dvir            ------------
Phum            ------------
Tcas            ------------
Nvit            ------------
Rpro            FKCMVWIPLSKL
```

Figure 11. Neurotactin Homologue Alignment. The yellow, highlighted regions indicate the transmembrane domain for each protein sequence.

Figure 12. Neurotactin Phylogenetic Tree



Figures 13a. – 13p. Nrt Homologue Hydropathy & Transmembrane Prediction Plots

Figure 13a. *Drosophila melanogaster*

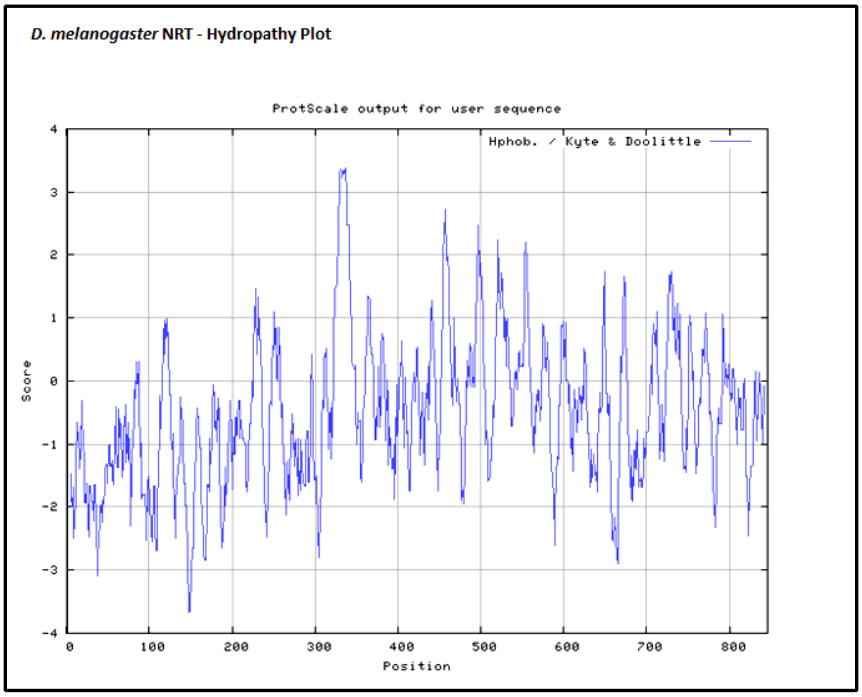


Figure 13a. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

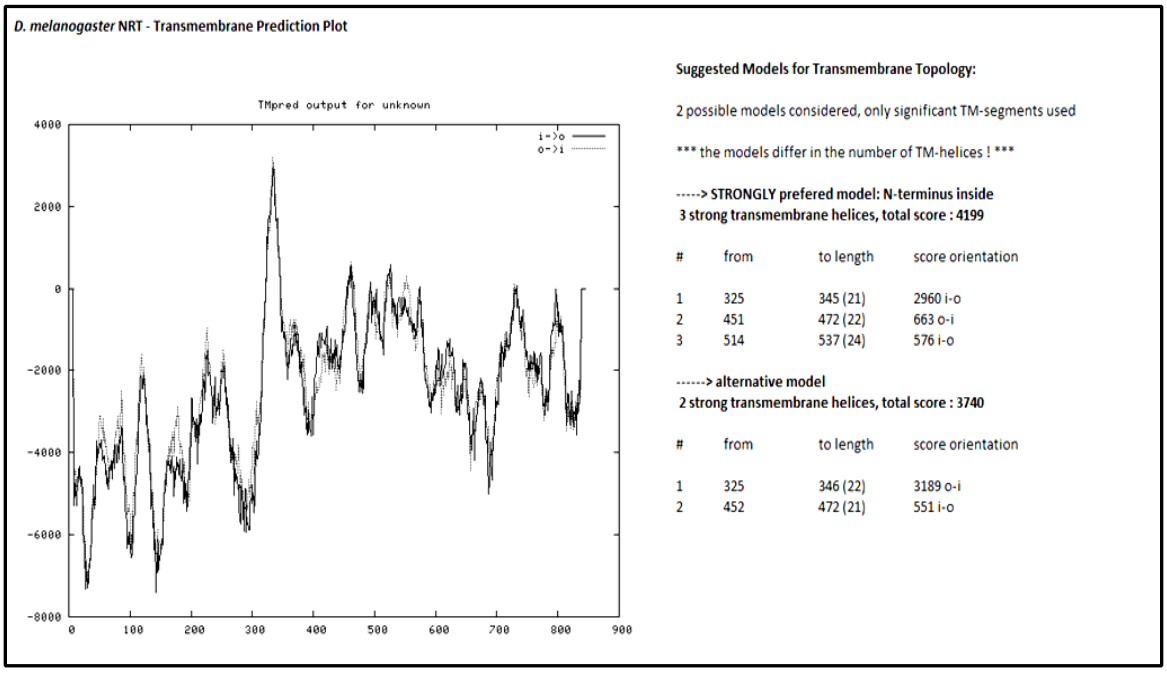


Figure 13b. *Drosophila ananassae*

***Drosophila ananassae* NRT- Hydropathy Plot**

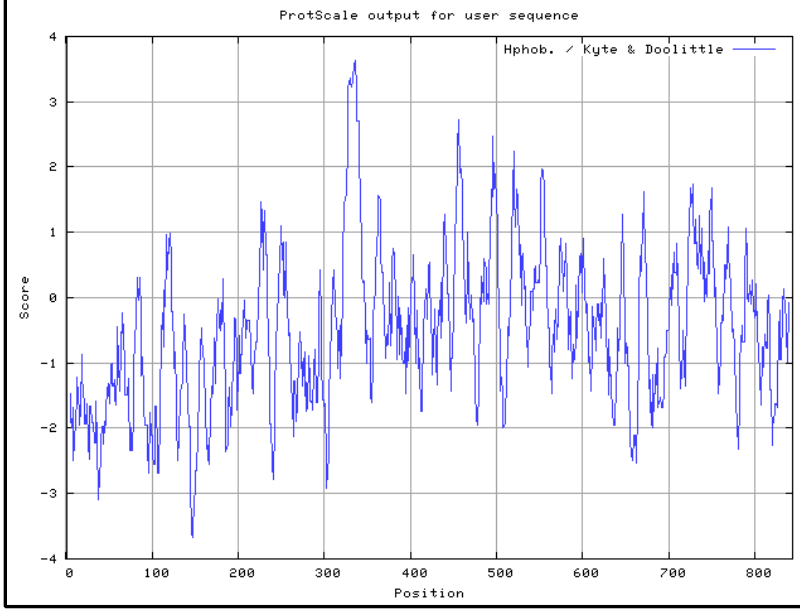
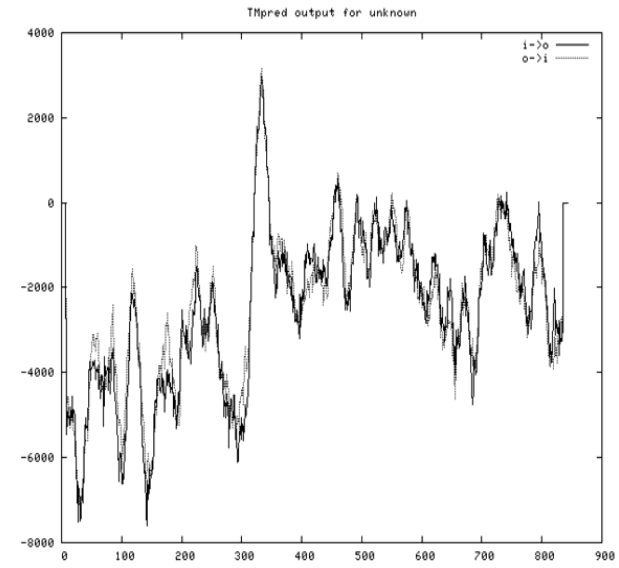


Figure 13b. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

***D. ananassae* NRT - Transmembrane Prediction**



Suggested models for transmembrane topology

2 possible models considered, only significant TM-segments used

-----> STRONGLY prefered model: N-terminus outside
2 strong transmembrane helices, total score : 3749

#	from	to length	score orientation
1	324	342 (19)	3181 o-i
2	448	471 (24)	568 i-o

------> alternative model
2 strong transmembrane helices, total score : 3658

#	from	to length	score orientation
1	324	344 (21)	2968 i-o
2	448	471 (24)	690 o-i

Figure 13c. *Drosophila grimshawi*

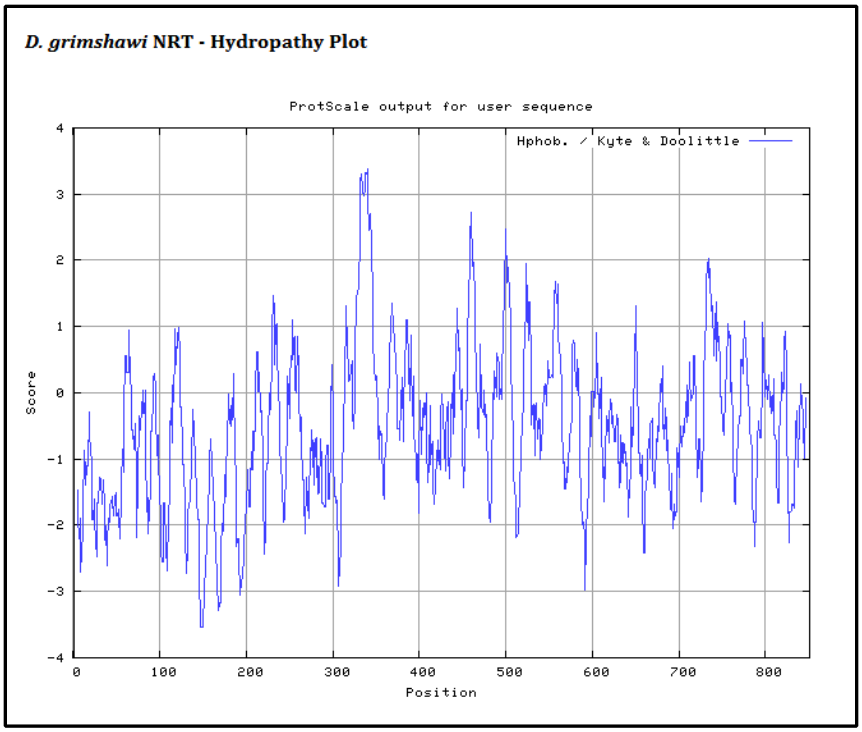


Figure 13c. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

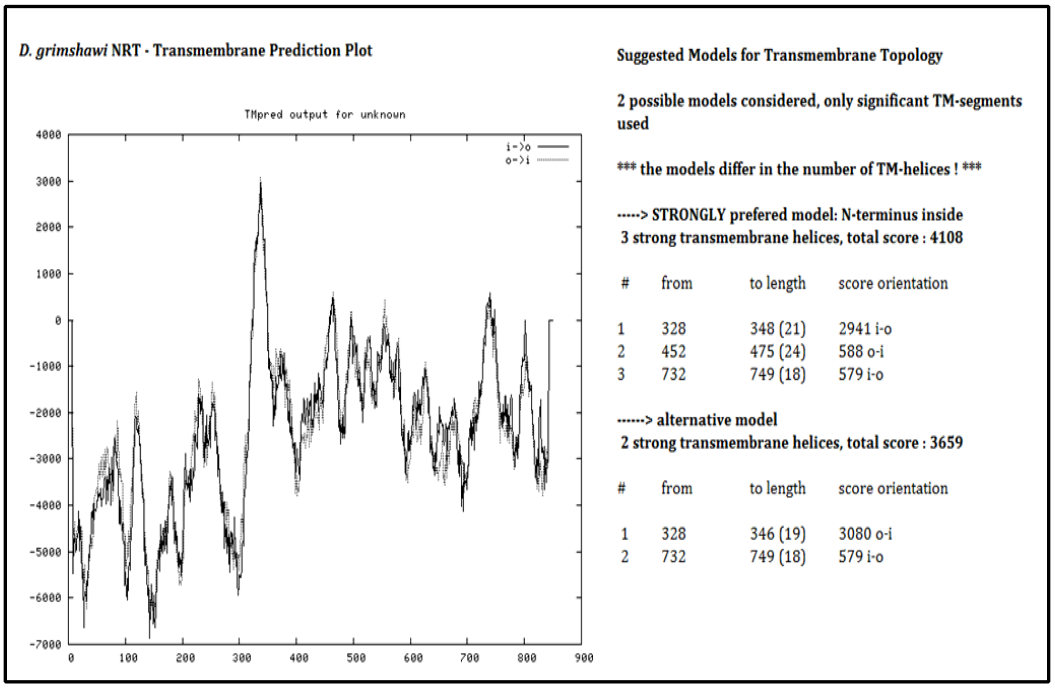


Figure 13d. *Drosophila pseudoobscura*

***D. pseudoobscura* NRT- Hydropathy Plot**

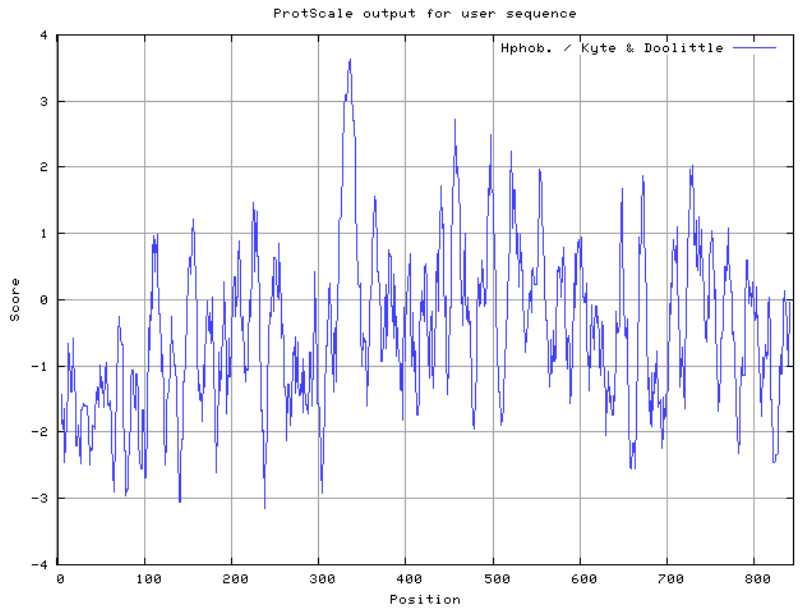
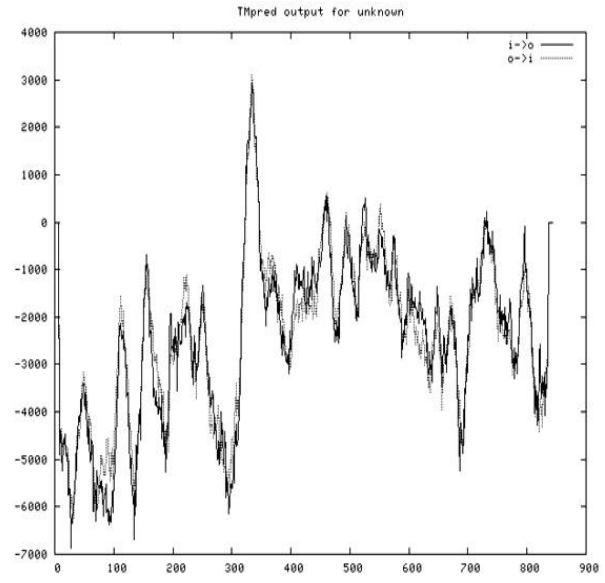


Figure 13d. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

***D. pseudoobscura* NRT - Transmembrane Prediction Plot**



Suggested Models for Transmembrane Topology:

2 possible models considered, only significant TM-segments used

-----> STRONGLY prefered model: N-terminus outside
2 strong transmembrane helices, total score : 3662

#	from	to length	score orientation
1	325	346 (22)	3110 o-i
2	449	472 (24)	552 i-o

------> alternative model
2 strong transmembrane helices, total score : 3527

#	from	to length	score orientation
1	325	345 (21)	2905 i-o
2	449	472 (24)	622 o-i

Figure 13e. *Drosophila virilis*

***D. virilis* NRT - Hydropathy Plot**

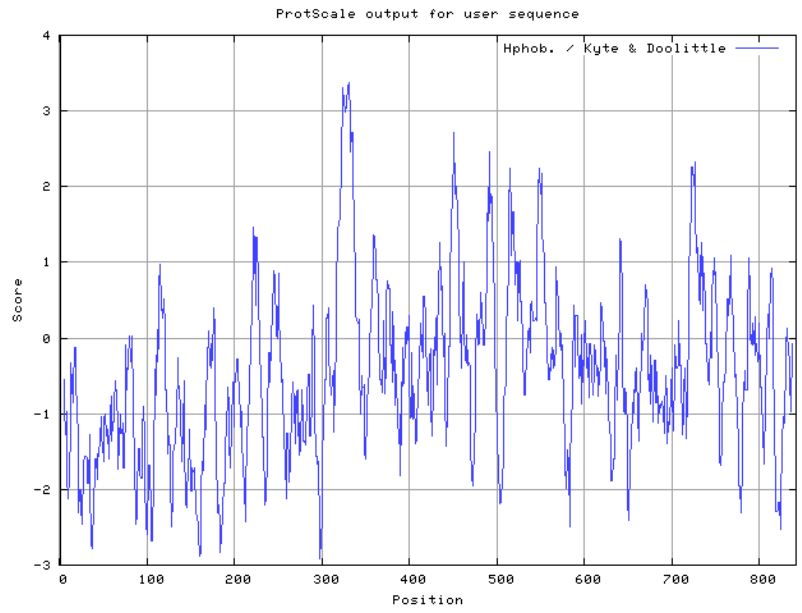
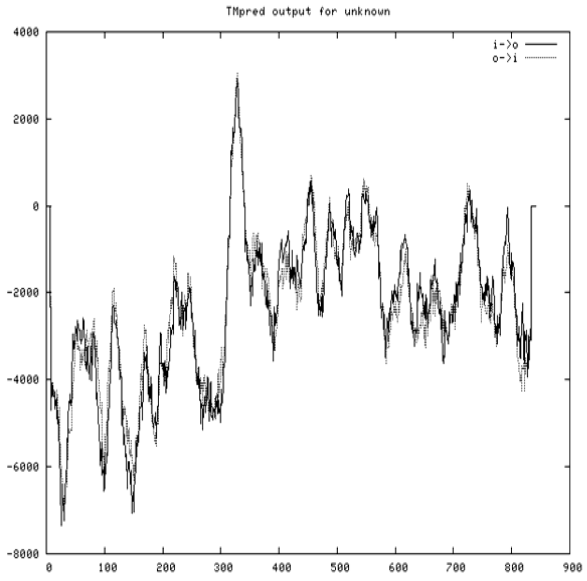


Figure 13e. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

***D. virilus* NRT - Transmembrane Prediction Plot**



Suggested Models for Transmembrane Topology:

2 possible models considered, only significant TM-segments used

------> STRONGLY prefered model: N-terminus outside
3 strong transmembrane helices, total score : 4296

#	from	to length	score orientation
1	319	337 (19)	3080 o-i
2	443	466 (24)	568 i-o
3	537	562 (26)	648 o-i

-------> alternative model
2 strong transmembrane helices, total score : 3631

#	from	to length	score orientation
1	319	339 (21)	2941 i-o
2	443	466 (24)	690 o-i

Figure 13f. *Drosophila willistoni*

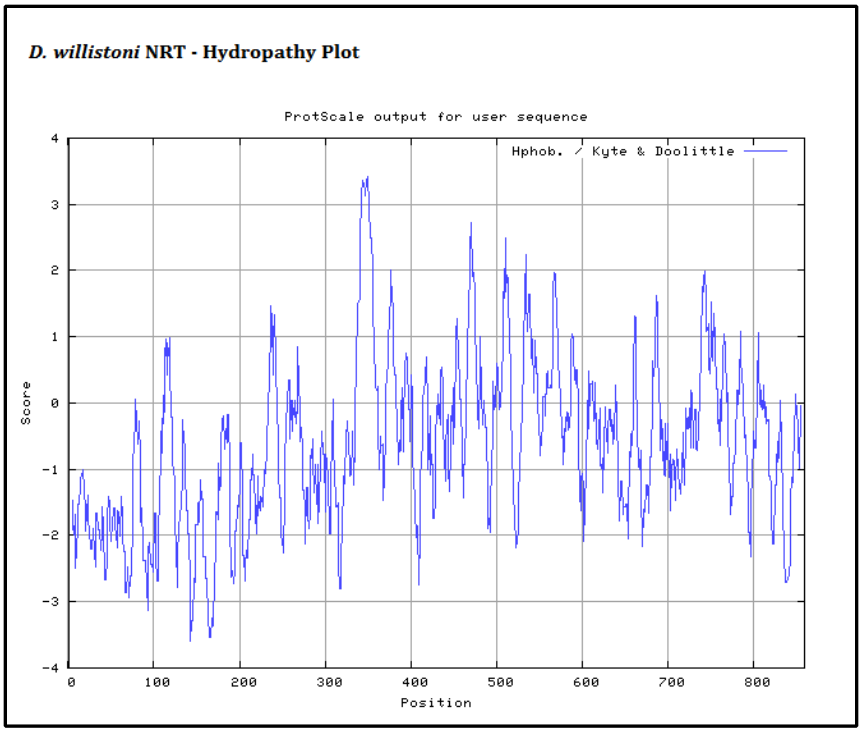


Figure 13f. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

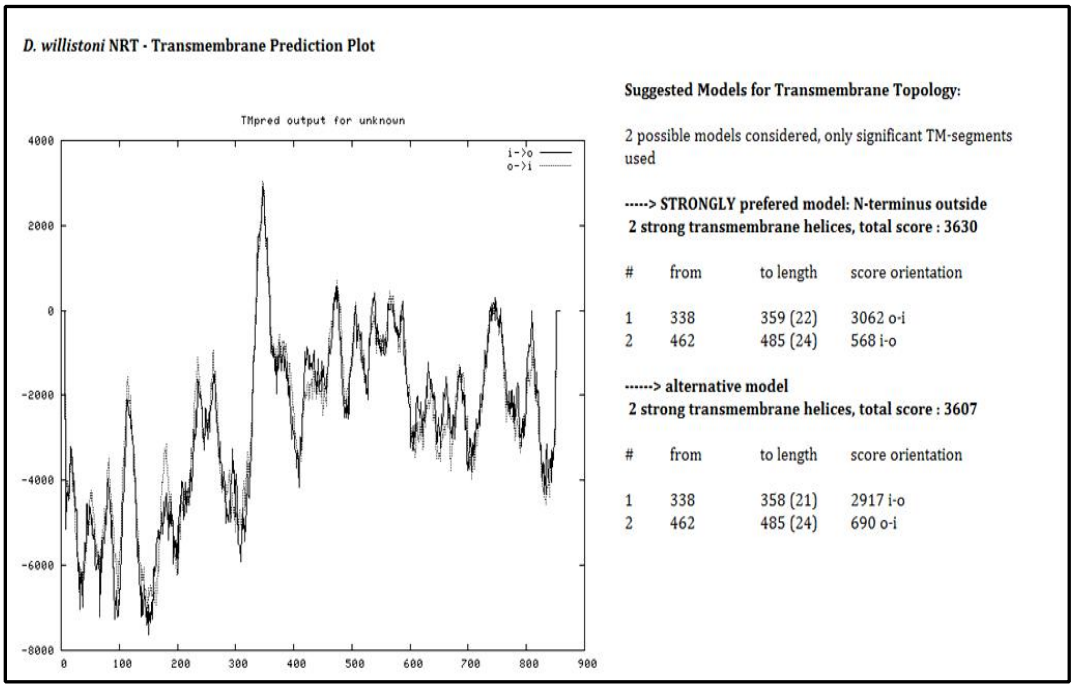


Figure 13g. *Ceratitis capitata*

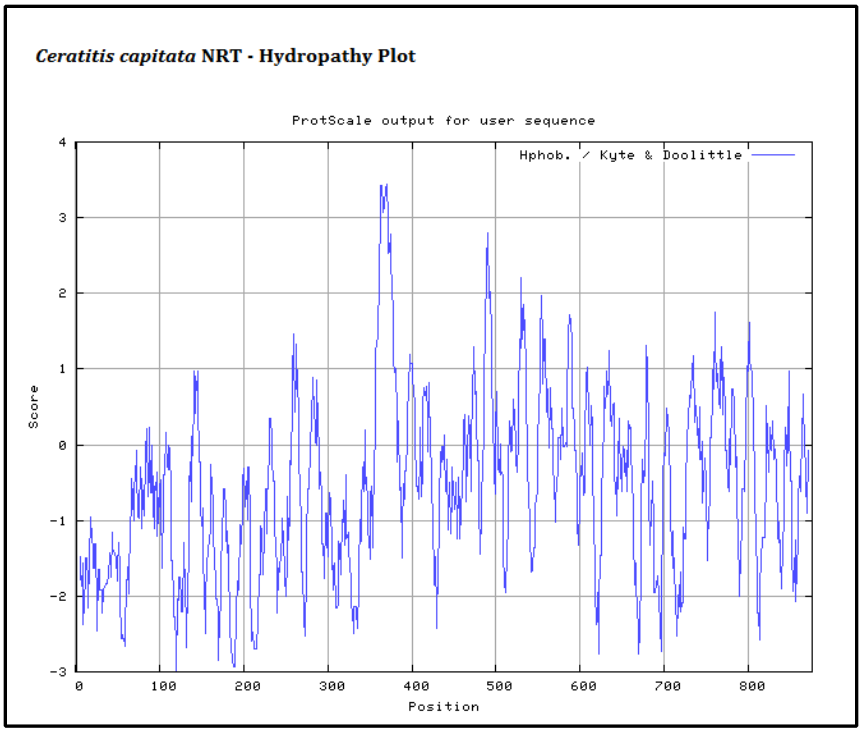


Figure 13g. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 350-400. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

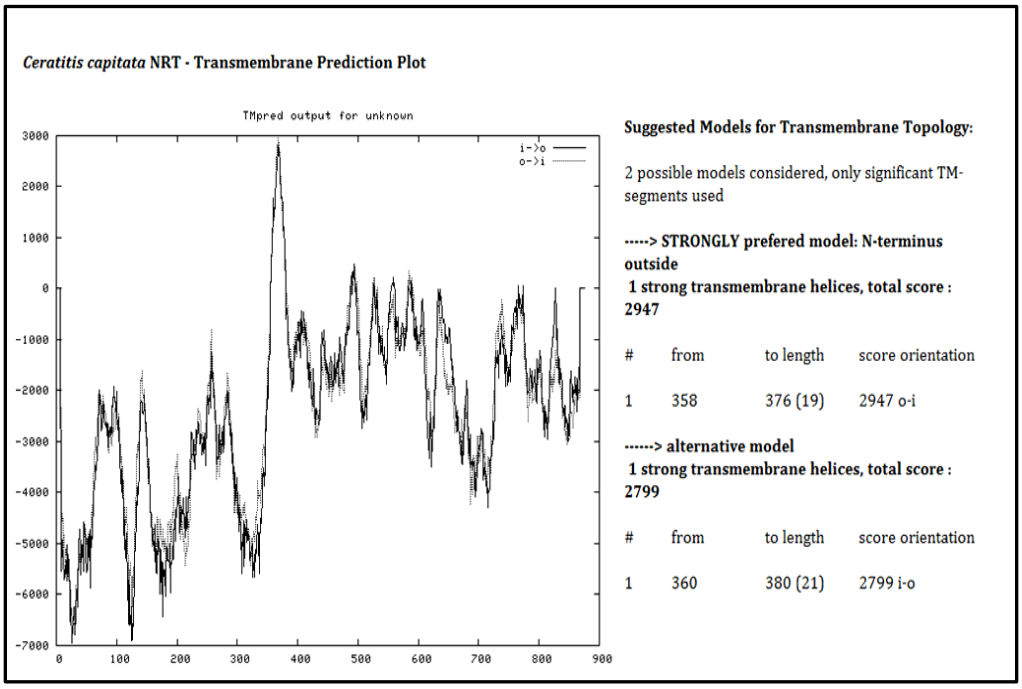


Figure 13h. *Musca domestica*

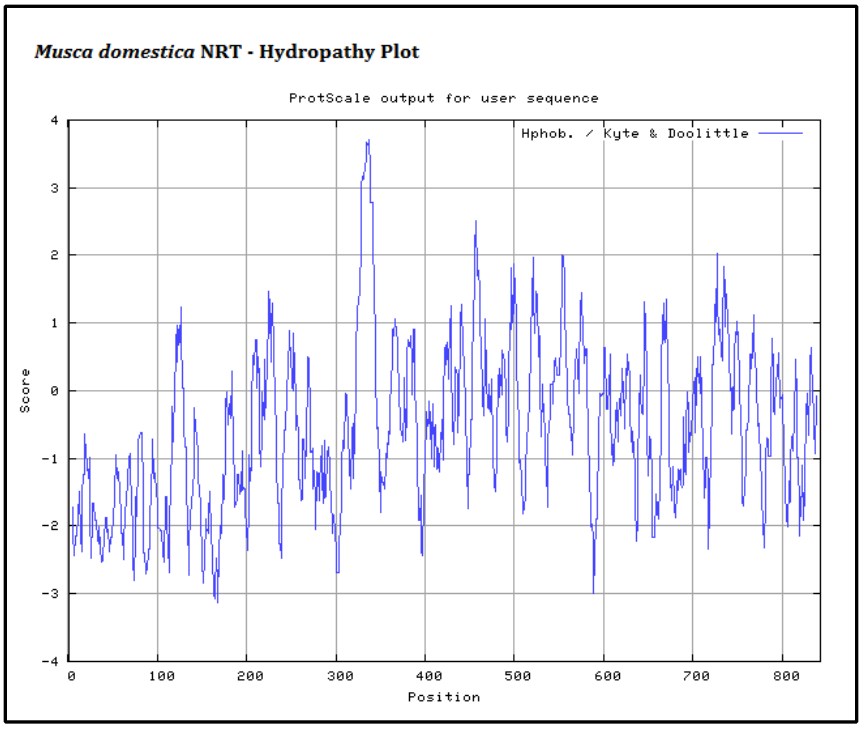


Figure 13h. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

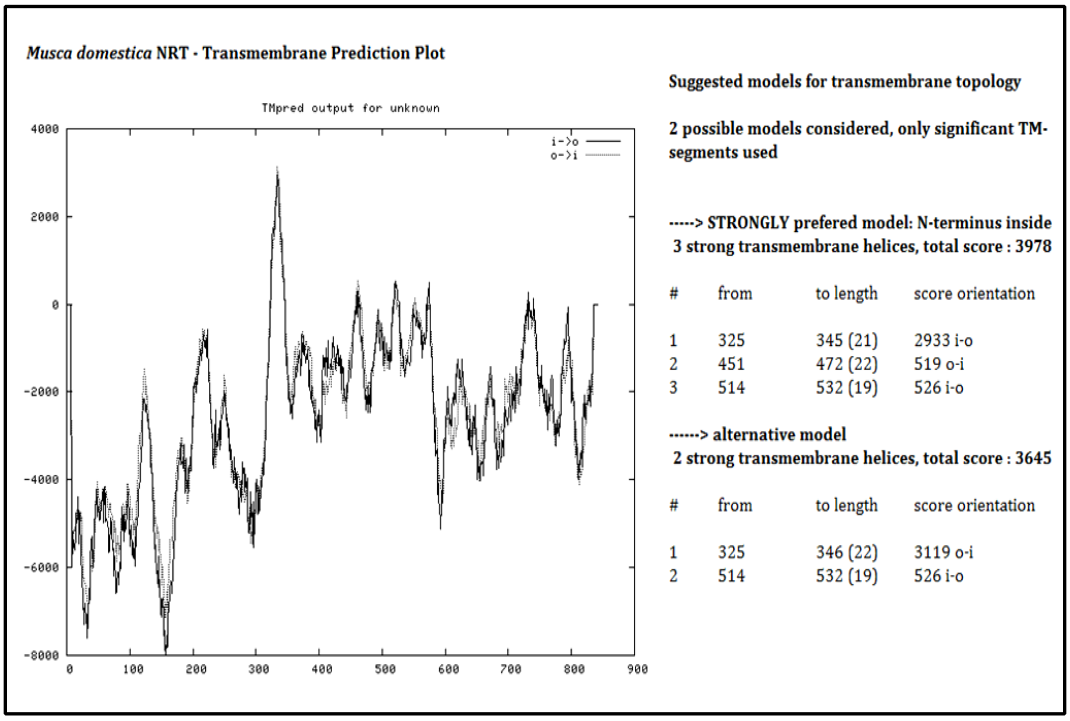


Figure 13i. *Glossina morsitans*

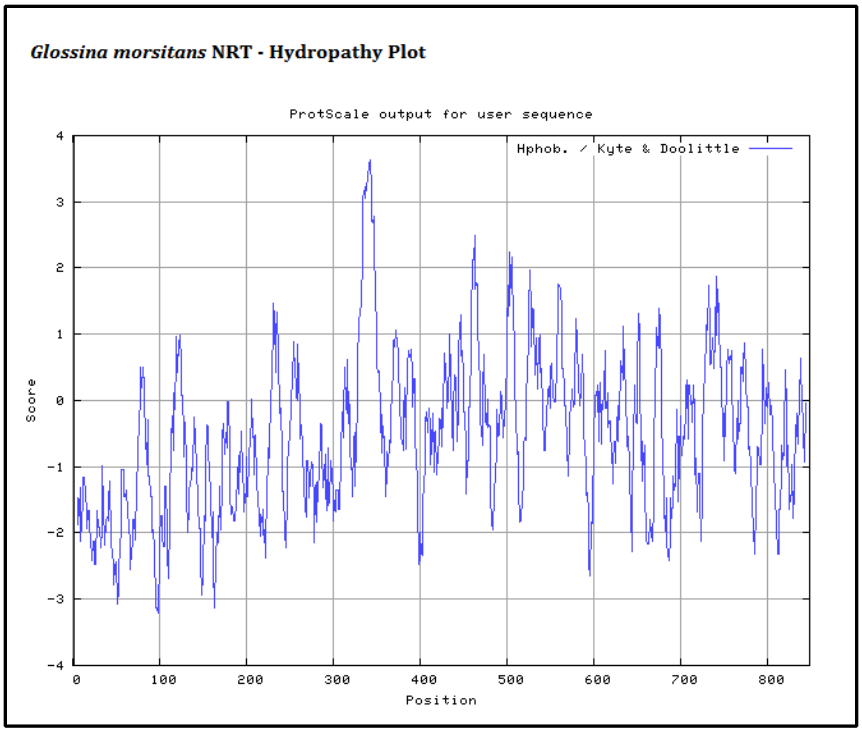


Figure 13i. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

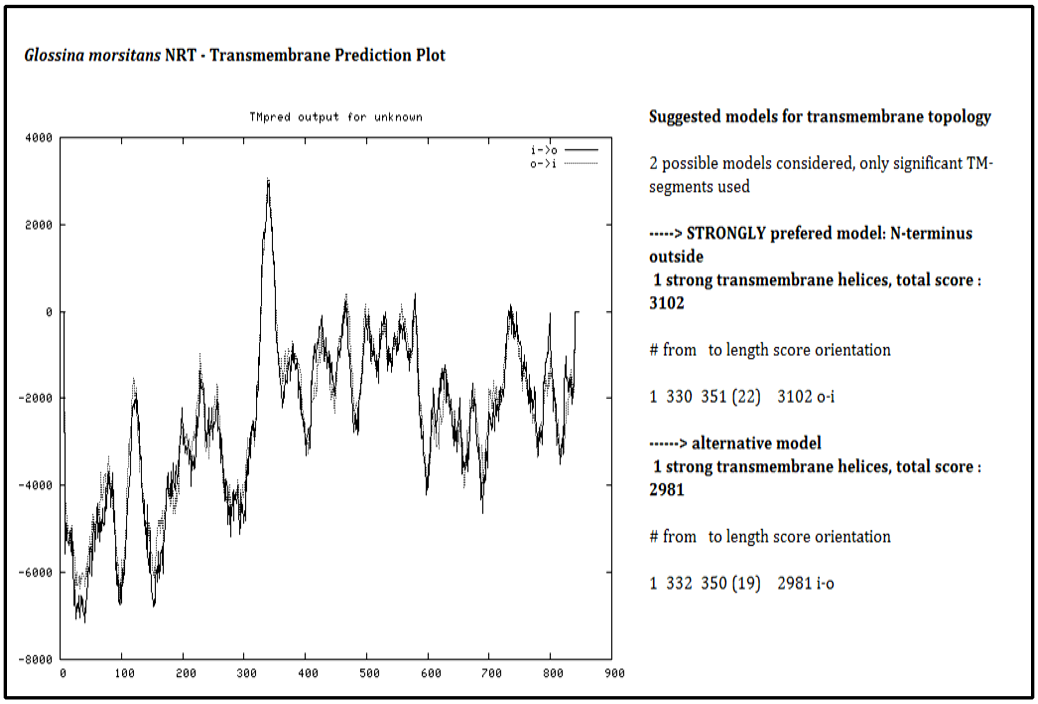


Figure 13j. *Aedes aegypti*

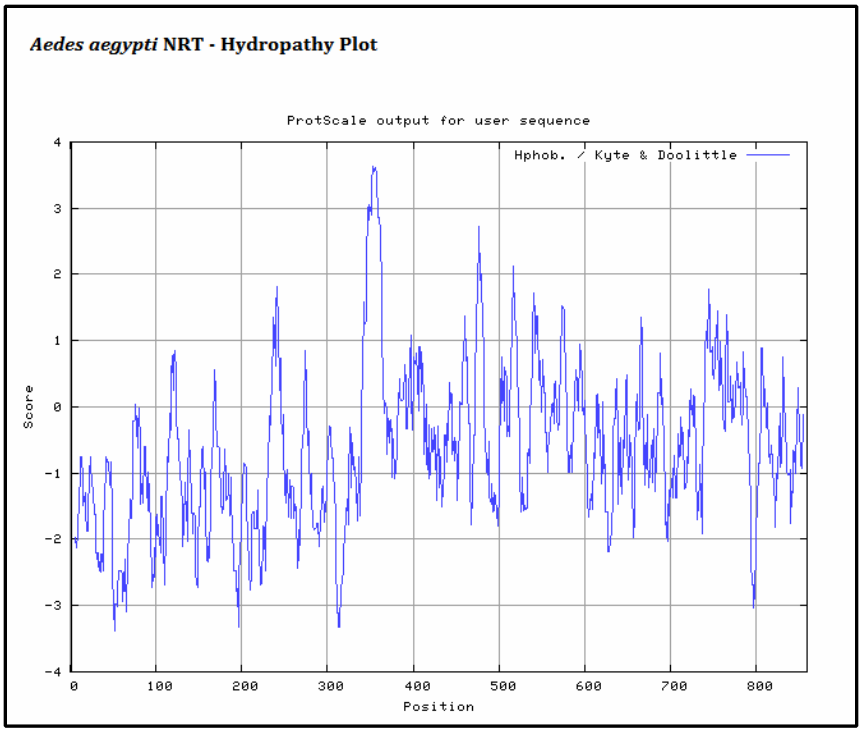


Figure 13j. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

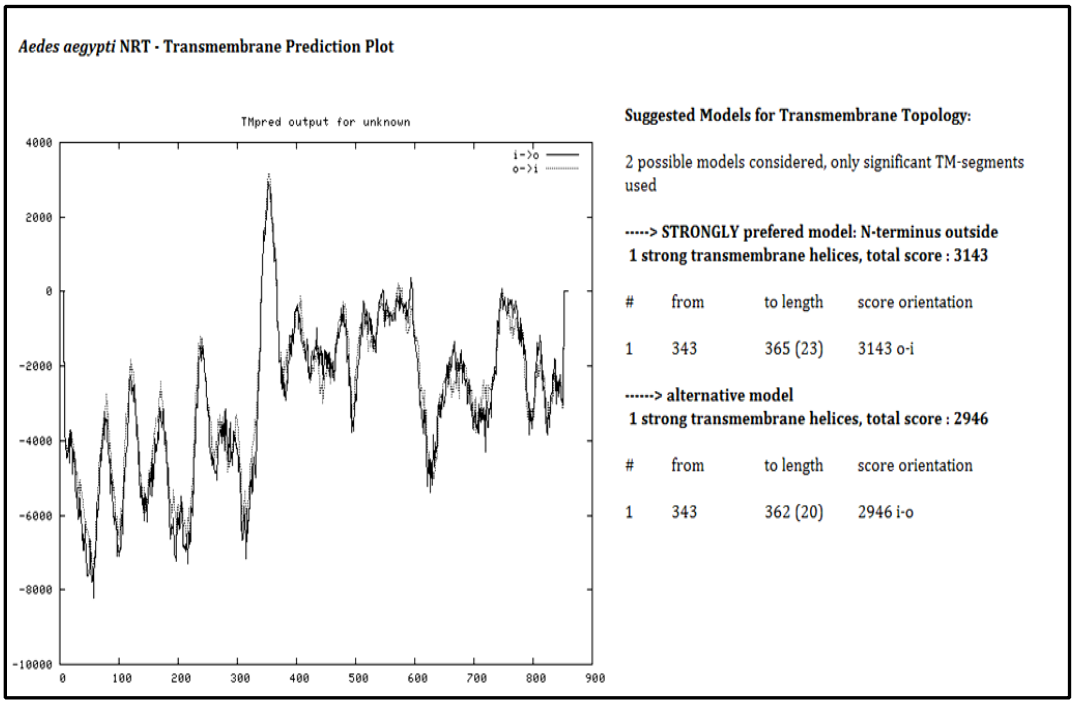


Figure 13k. *Anopheles gambiae*

***Anopheles gambiae* NRT - Hydropathy Plot**

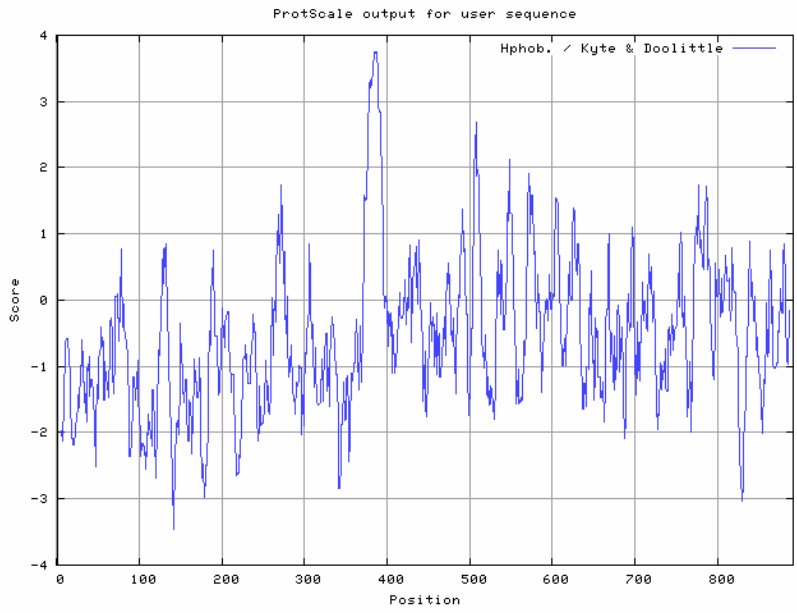
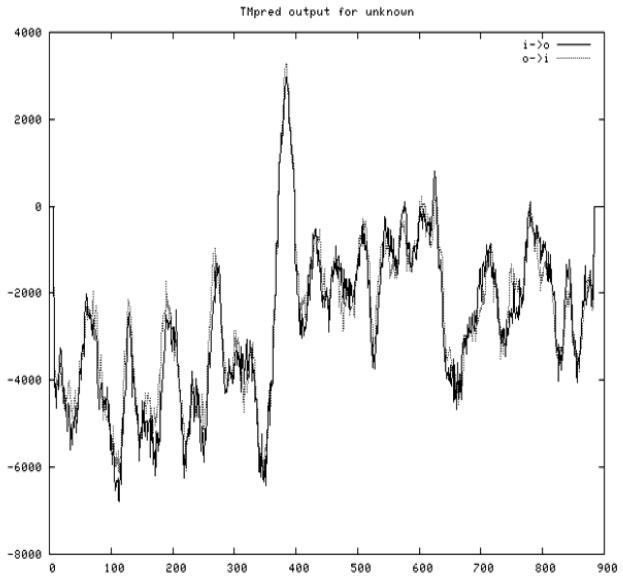


Figure 13k. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 350-400. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

***Anopheles gambiae* NRT - Transmembrane Prediction Plot**



Suggested Models for Transmembrane Topology:

2 possible models considered, only significant TM-segments used

-----> STRONGLY prefered model: N-terminus outside
2 strong transmembrane helices, total score : 4098

#	from	to length	score orientation
1	374	393 (20)	3287 o-i
2	617	636 (20)	811 i-o

------> alternative model
1 strong transmembrane helices, total score : 2961

#	from	to length	score orientation
1	374	396 (23)	2961 i-o

Figure 13l. *Culex quinquefasciatus*

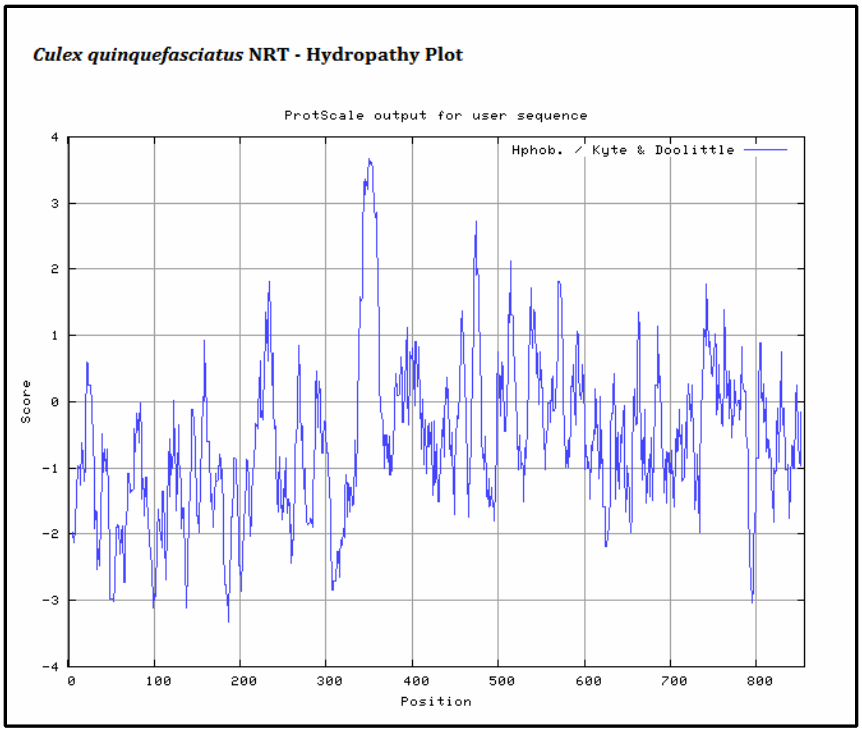


Figure 13l. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

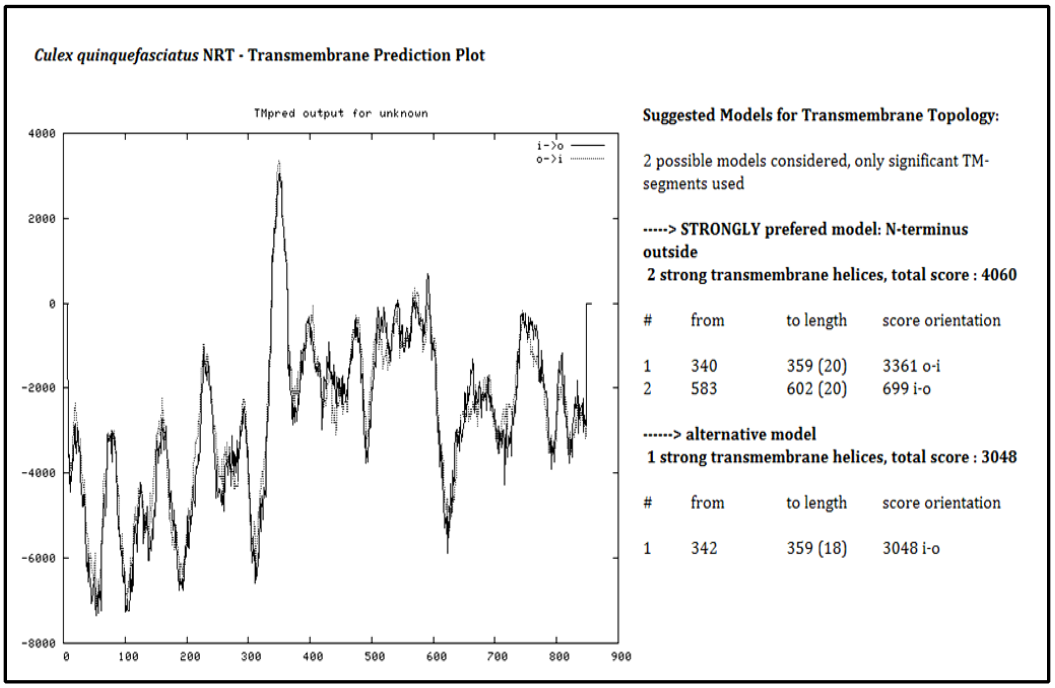


Figure 13m. *Tribolium castaneum*

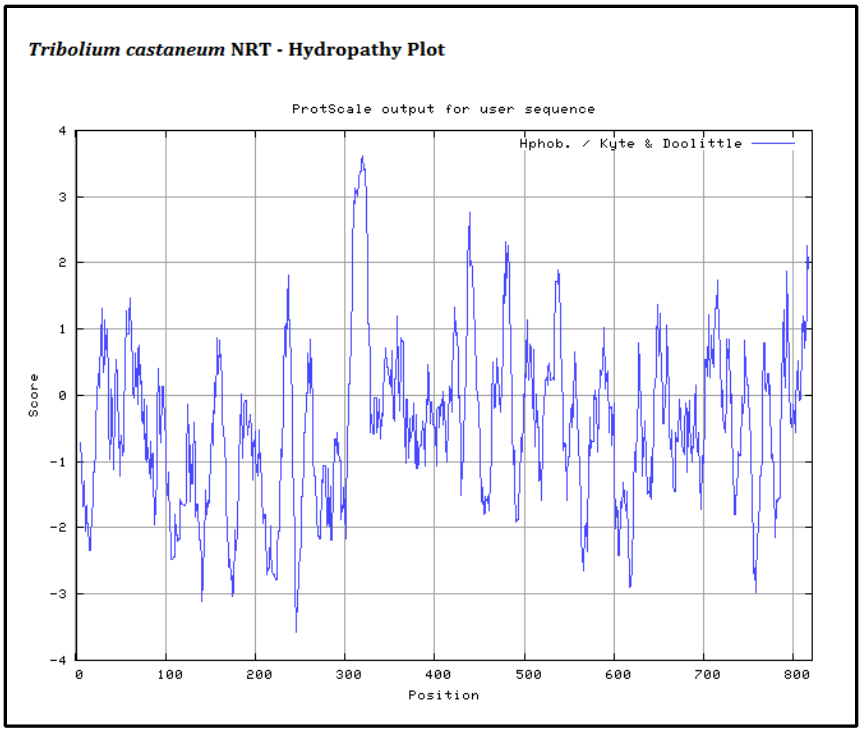


Figure 13m. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

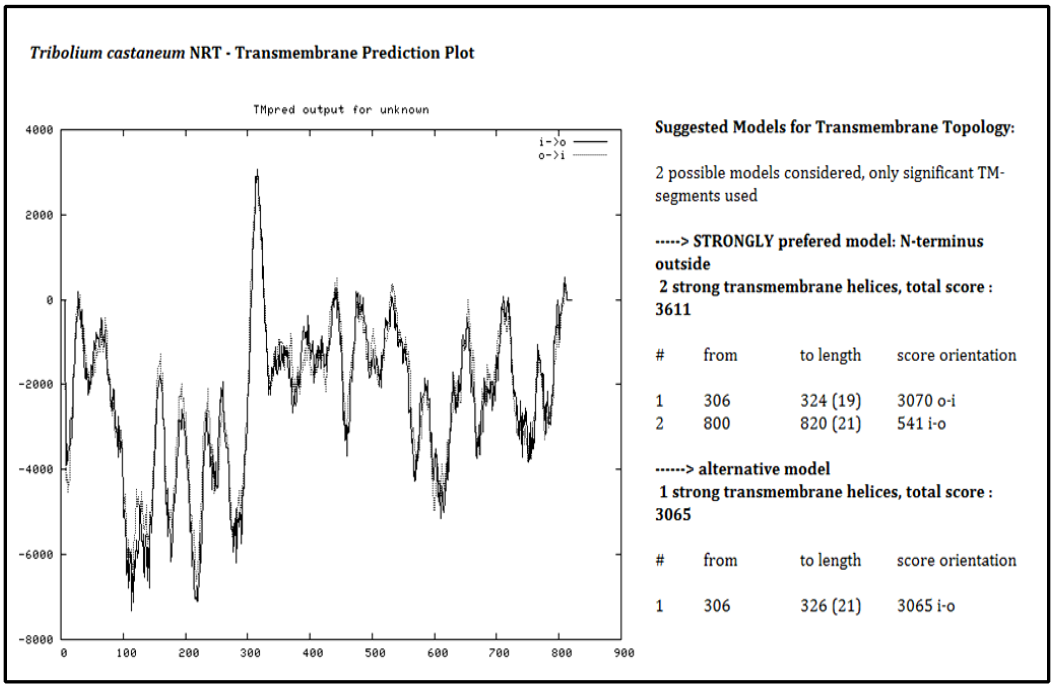


Figure 13n. *Nasonia vitripennis*

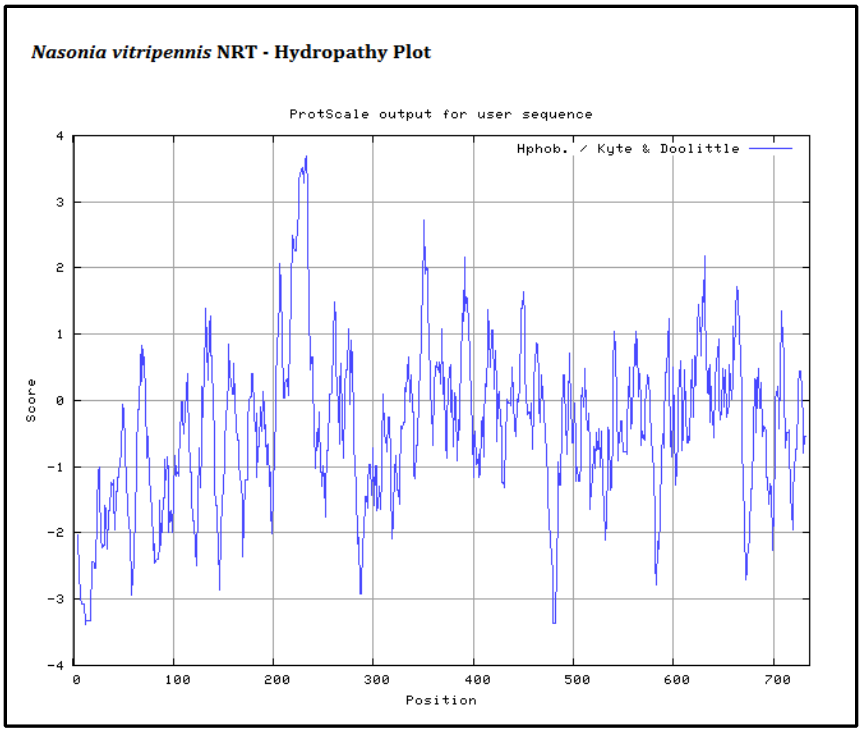


Figure 13n. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 200-250. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

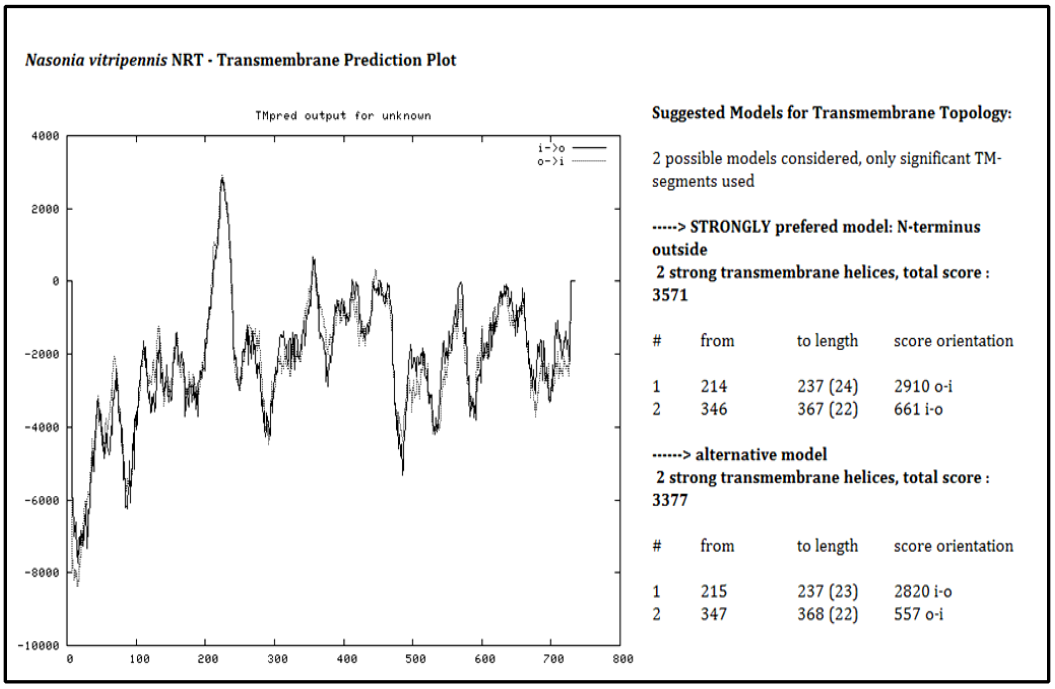


Figure 13o. *Pediculus humanus*

***Pediculus humanus* NRT - Hydropathy Plot**

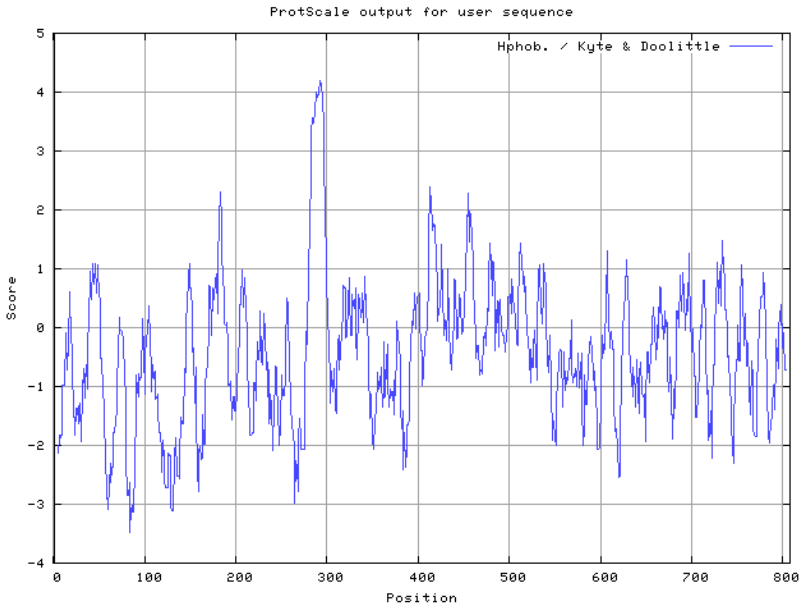
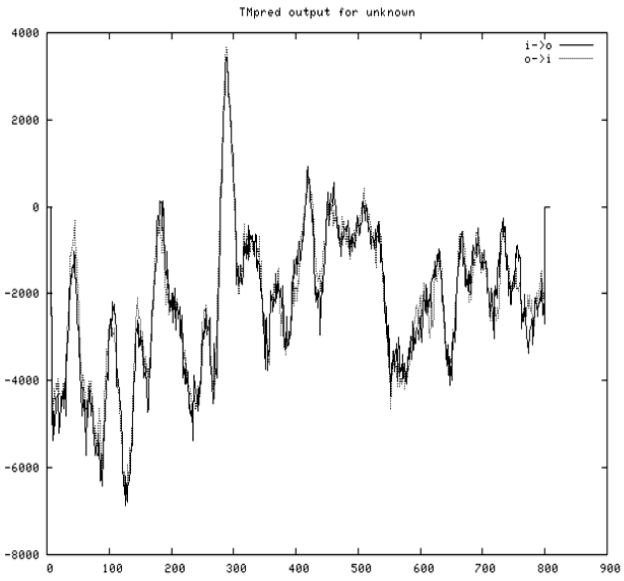


Figure 13o. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 250-300. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins

***Pediculus humanus* NRT - Transmembrane Prediction Plot**



Suggested Models for Transmembrane Topology:

2 possible models considered, only significant TM-segments used

----> STRONGLY prefered model: N-terminus inside
3 strong transmembrane helices, total score : 4658

#	from	to length	score orientation
1	281	298 (18)	3438 i-o
2	410	430 (21)	670 o-i
3	451	472 (22)	550 i-o

------> alternative model
2 strong transmembrane helices, total score : 4609

#	from	to length	score orientation
1	279	299 (21)	3676 o-i
2	410	430 (21)	933 i-o

Figure 13p. *Rhodnius prolixus*

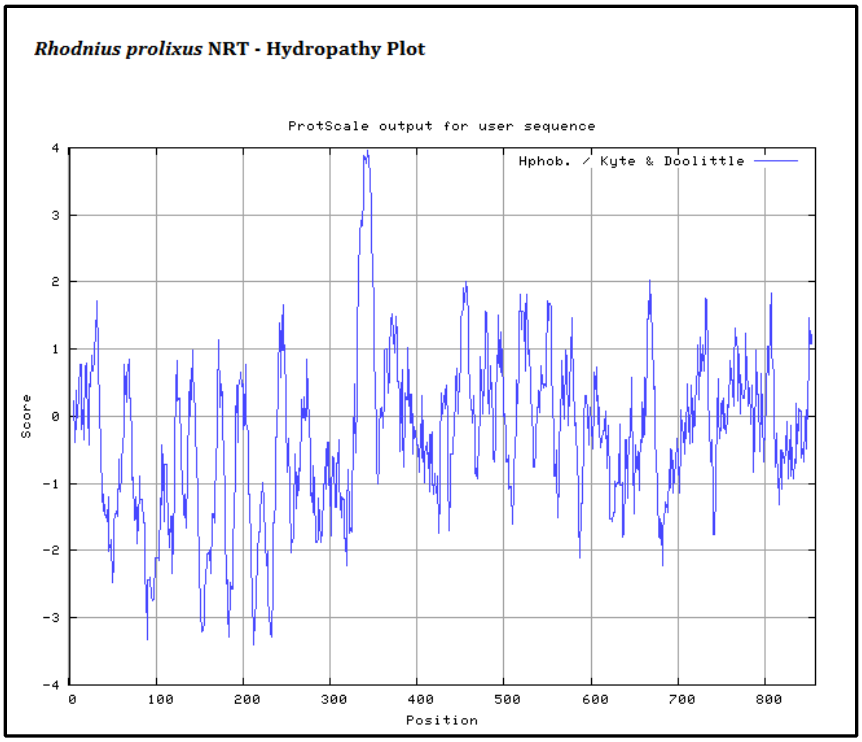
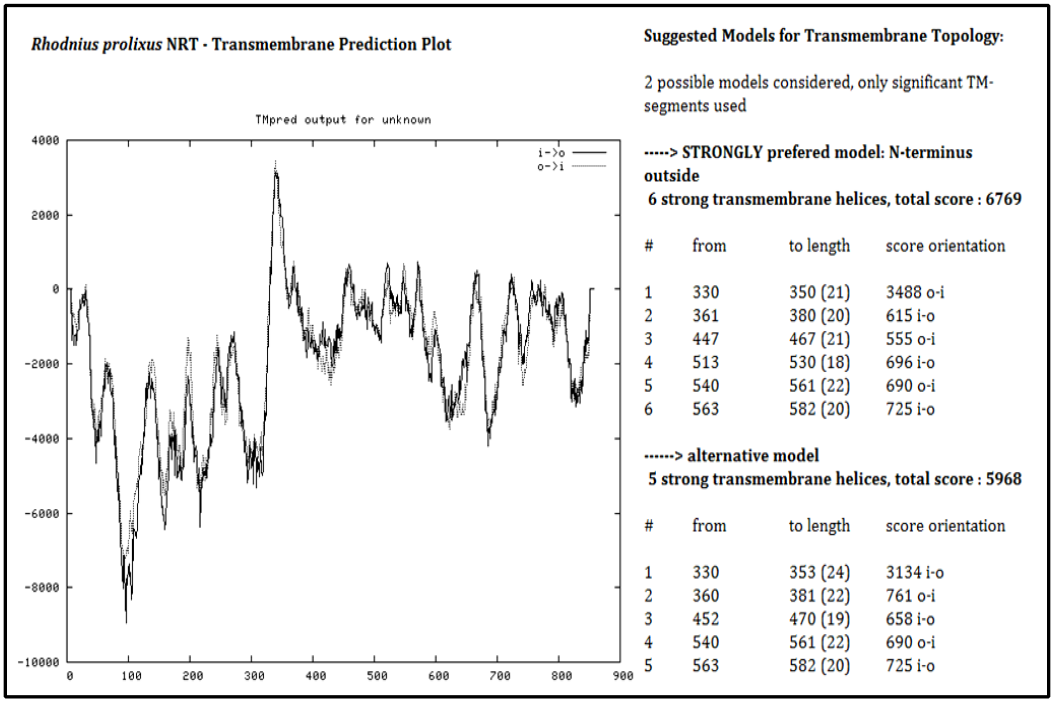


Figure 13p. Hydropathy plot on the Left and Transmembrane Prediction Plot Below. The Hydropathy plot depicts a strong hydrophobic peak between 300-350. This is also the same region where the Transmembrane Prediction Plot suggests the transmembrane domain of Nrt exists. Note in the Hydropathy plot there is no signal sequence hydrophobic peak- a feature of type II transmembrane proteins



Suggested Models for Transmembrane Topology:

2 possible models considered, only significant TM-segments used

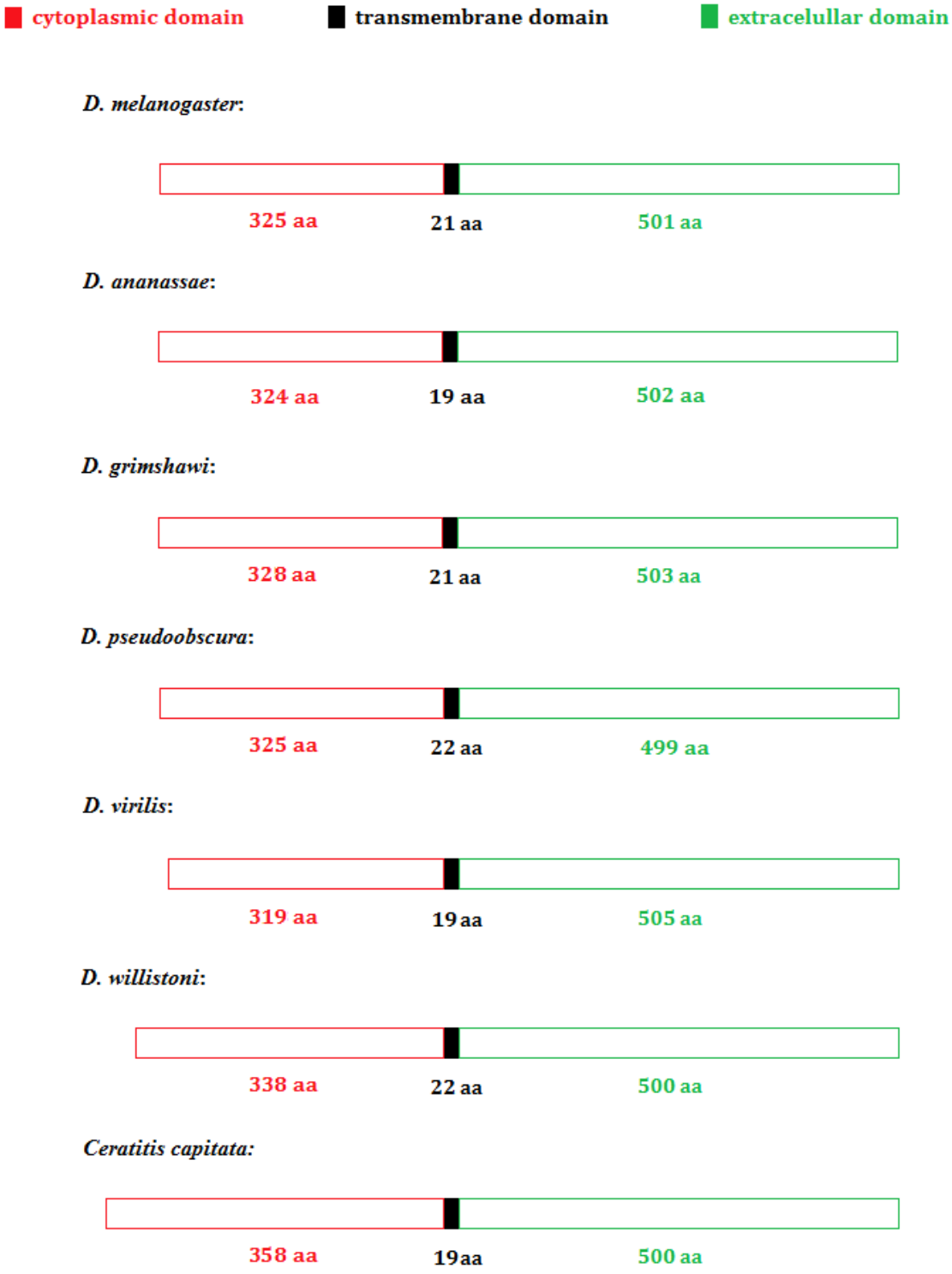
-----> STRONGLY prefered model: N-terminus outside
6 strong transmembrane helices, total score : 6769

#	from	to length	score orientation
1	330	350 (21)	3488 o-i
2	361	380 (20)	615 i-o
3	447	467 (21)	555 o-i
4	513	530 (18)	696 i-o
5	540	561 (22)	690 o-i
6	563	582 (20)	725 i-o

------> alternative model
5 strong transmembrane helices, total score : 5968

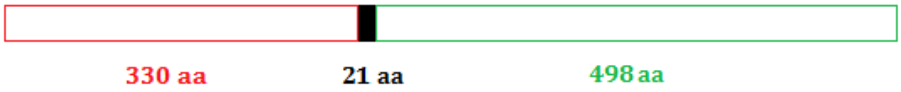
#	from	to length	score orientation
1	330	353 (24)	3134 i-o
2	360	381 (22)	761 o-i
3	452	470 (19)	658 i-o
4	540	561 (22)	690 o-i
5	563	582 (20)	725 i-o

Figure 14. Neurotactin Homologues

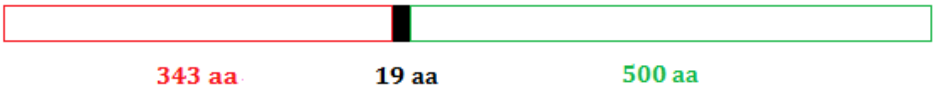


***Musca domestica*:**

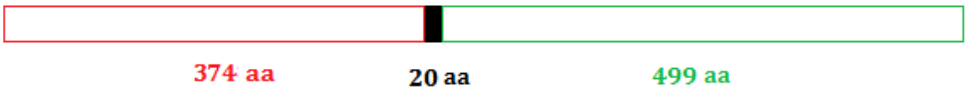
***Glossina morsitans*:**



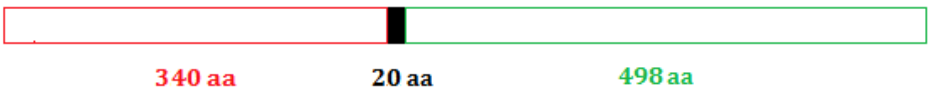
***Aedes aegypti*:**



***Anopheles gambiae*:**



***Culex quinquefasciatus*:**



***Tribolium castaneum*:**

***Nasonia vitripennis*:**

***Pediculus humanus*:**

***Rhodnius prolixus*:**

REFERENCES:

[1]De la Escalera, Santiago, et al. "Characterization and Gene Cloning of Neurotactin, a *Drosophila* Transmembrane Protein Related to Cholinesterases." *EMBO Journal* 9.11(1990): 3593-601. Print.

[2]Hortsch, Michael, et al. "*Drosophila* Neurotactin, a Surface Glycoprotein with Homology to Serine Esterases, Is Dynamically Expressed during Embryogenesis." *Development* 110 (1990): 1327-40.Print.

[3]Speicher, Stephan, et al. "Neurotactin Functions in Concert with Other Identified CAMs in Growth Cone Guidance in *Drosophila*." *Neuron* 20 (1998): 221-33. Print.

[4]Gilbert, Mary M., and Vanessa J. Auld. "Evolution of CLAMS (Cholinesterase-Like Adhesion Molecules): Structure and Function during Development." *Frontiers in Bioscience* 10 (2005): 2177-92. Print.

[5]Darboux, I., et al. "The Structure-Function Relationships in *Drosophila* Neurotactin Shows That Cholinesterasic Domains May Have Adhesive Properties." *EMBO Journal* 15.18 (1996): 4835-43. Print.

[6]Frémion, F., et al. "Amalgam Is a Ligand for the Transmembrane Receptor Neurotactin and Is Required for Neurotactin-Mediated Cell Adhesion and Axon Fasciculation in *Drosophila*." *EMBO Journal* 19.17 (2000): 4463-72. Print.

[7] Barthalay, Yves., et al. "*Drosophila* Neurotactin Mediates Heterophilic Cell Adhesion." *EMBO Journal* 9.11 (1990): 3603-09. Print.

[8]Liebl, Eric C., et al. "Interactions between the Secreted Protein Amalgam, Its Transmembrane Receptor Neurotactin and the Abelson Tyrosine Kinase Affect Axon Pathfinding." *Development* 130 (2003): 3217-26. Print.

[9]Seeger, M. A., L. Haffley, and T. C. Kaufman. "Characterization of Amalgam: A Member of the Immunoglobulin Superfamily from *Drosophila*." *Cell* 55 (1988): 589-600. Print.